

Ordering in voter models on networks: Exact reduction to a single-coordinate diffusion

R. A. Blythe

SUPA, School of Physics and Astronomy, University of Edinburgh, Mayfield Road,
Edinburgh EH9 3JZ, UK

9 August 2010

Abstract. We study the voter model and related random-copying processes on arbitrarily complex network structures. Through a representation of the dynamics as a particle reaction process, we show that a quantity measuring the degree of order in a finite system is, under certain conditions, exactly governed by a universal diffusion equation. Whenever this reduction occurs, the details of the network structure and random-copying process affect only a single parameter in the diffusion equation. The validity of the reduction can be established with considerably less information than one might expect: it suffices to know just two characteristic timescales within the dynamics of a single pair of reacting particles. We develop methods to identify these timescales, and apply them to deterministic and random network structures. We focus in particular on how the ordering time is affected by degree correlations, since such effects are hard to access by existing theoretical approaches.

1. Introduction

The voter model [1–4] is arguably the simplest nonequilibrium ordering process. Its dynamics are perhaps most easily described within its original formulation as a model for population dynamics [1]. Each site of a lattice (or network) is occupied by an individual belonging to one of a number of species. In each timestep, a site is chosen at random, and the individual at that site is replaced by an offspring (copy) of an individual selected at random from a neighbouring site. In a finite system, one of the possible absorbing states—in which only a single species remains—is eventually reached. From a physical perspective, the interest in this process is that order is arrived at purely through fluctuations, rather than driven by surface tension as in the Ising model (see e.g., [5, 6]; also [7] for a recent review).

Many variants of the basic voter model have been investigated, some quite recently: examples include those where multiple individuals occupy a site [8], where the site chosen in each timestep contains the individual that is copied rather than replaced [9, 10], or where different sites have different probabilities of being chosen for update [11]. Specific instances drawn from this wider class of *random-copying* models have applications in population genetics, ecology and cultural evolution, as reviewed in [12]. Despite their

simplicity, these models reproduce observed patterns of species diversity in ecosystems [13] and of the cultural dynamics of features like baby names [14]. They also play an important role as *null models* of evolution in the absence of selection [15].

A stochastic equation of motion for the frequency ξ of a chosen species is straightforwardly obtained in the mean-field limit (i.e., when the system lacks spatial structure). More precisely, when each site is visited for update at a uniform rate m , and the individual landing on the target site is sampled uniformly from the entire population, the probability distribution of ξ is—for sufficiently large systems—governed by

$$\frac{\partial}{\partial t}P(\xi, t) = \frac{m}{N_e} \frac{\partial^2}{\partial \xi^2} \xi(1 - \xi)P(\xi, t). \quad (1)$$

Derivations of this equation can be found in many places, e.g., [12, 16, 17]. These stochastic dynamics have three key properties. First, there is no deterministic (drift) term in the above equation, confirming that order is attained through fluctuations alone. Second, all timescales of the ordering process are proportional to the parameter N_e : in the mean field, this parameter is simply equal to the number of sites, N . Finally, the full, time-dependent solution for $P(\xi, t)$ from arbitrary initial condition is known [16]. Then, for example, the probability that complete order has been attained by time t is given by the combination $P(0, t) + P(1, t)$.

Many recent works have switched focus to models defined on nontrivial network structures, notably *heterogeneous* networks in which the degree can vary widely between nodes [9, 10, 18–27]. Such networks are particularly relevant to cultural applications, where the interaction frequency between agents need not be purely a function of spatial proximity. Remarkably, it has been found that even in the presence of significant heterogeneity, the ordering dynamics is well-described by the mean-field equation (1). For example, the mean time to reach complete order as predicted by (1) has been found to apply in models with complex spatial structure [10, 23–27]. The only differences are: (i) that ξ is a weighted average of a species frequency over the network; and (ii) the *effective size*, N_e , need not equal the actual network size N (or even scale linearly with it). Thus, a good understanding of ordering behaviour within random-copying models on networks can be gained if one can identify when a reduction to (1) is possible, and if so, what the value of N_e is. For example, the entire distribution of ordering times can be read off from the exact solution of (1).

In this work we will demonstrate that the mean-field diffusion equation (1) is in fact *exact* for general random-copying processes on arbitrary networks at all but the very earliest times when there is a separation of emergent timescales within a two-particle coalescence reaction dynamics that is *dual* to the random-copying process [28, 29]. In several earlier works [9, 10, 23–27], the equation (1) was typically obtained by making specific assumptions and approximations, and the fundamental physical reason why the simple diffusion (1) should apply was not clearly identified.

The central object in our analysis is the *quasi-stationary* distribution of a pair of particles that have survived for a long time without reacting. This state decays at some characteristic rate, and if its lifetime greatly exceeds the relaxation time to the

quasi-stationary state from a given initial condition, the model exhibits the separation of timescales that is required for the validity of (1). The effective size N_e is then proportional to the single, dominant timescale in the reaction dynamics, that is, the lifetime of the quasi-stationary state.

This paper is divided into two parts. In the first part (Sections 2–5) we define the general class of random-copying models that includes well-studied instances, like the voter model, as special cases, and derive our main results. These are: Eq. (1), the condition for its validity and the expression for the fundamental parameter N_e . As we will need to switch repeatedly between the forward-time formulation of the dynamics in terms of a diffusion equation and an equivalent backward-time particle reaction process whose details are not widely discussed in the physics literature, we must unfortunately contend with lengthy, but necessary, preliminaries in Sections 2–4. Once this is done, the derivations of our main results follow in Section 5. We emphasise that the validity of Eq. (1) can be determined by considering only a single pair of coalescing random walkers on the network. In principle, ascertaining whether the system is completely ordered requires knowledge of the state of all N sites simultaneously: hence being able to determine whether the considerably simpler single-coordinate diffusion applies without having to solve the full dynamics is of great practical benefit.

In the second part of this paper, we go beyond purely formal statements about random-copying dynamics and focus on strategies to estimate the emergent timescales of the two-particle reaction process. These methods are developed in Section 6 and used to demonstrate the validity of the diffusion (1) for specific combinations of network structures and copying dynamics in Section 7. Although these methods are restricted to random-copying processes that map onto time-reversal symmetric random walks, it turns out that the most widely-studied variations on voter-model dynamics [9, 10, 20, 26, 27] fall into this category. On the other hand, we do not (in contrast to some existing treatments [9, 10, 23, 25–27]) invoke a coarse-graining over multiple sites of the network to obtain a diffusion equation. Thus, we can use our analytical methods to investigate the effects of degree correlations on the effective size N_e that appears in Equation (1), which until now has been accessible only by computer simulation.

We remark that the idea that random-copying dynamics in the presence of spatial structure may be adequately represented by a mean-field model with some effective size N_e has a very long history in population genetics, dating back to the work of Wright from the 1930s [30]. Furthermore, various separation of timescales arguments have been employed as a means to identify an effective size [31–36]. In this work, our main contribution to this line of enquiry is to highlight the two particular emergent timescales that are relevant in determining the applicability (or otherwise) of a single effective size: their importance does not appear to have been fully recognised before. We furthermore show that this separation exists even in cases where it has been argued that there will be more than one relevant timescale in the dynamics, and that a description in terms of a single effective size will not apply [35]. We discuss these points in much more detail in the conclusion, Section 8, along with a number of remaining open questions.

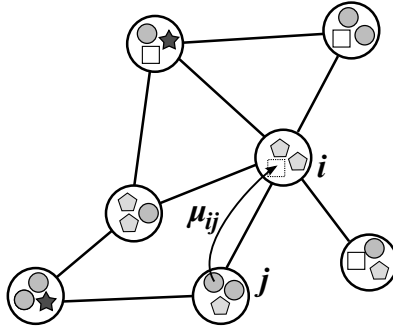


Figure 1. Random-copying dynamics on a network. Here, each site hosts $M = 3$ individuals and there are four species in total. The individual with dotted outline on site i has been selected for replacement by copying an individual on site j . This event takes place as a Poisson process with rate μ_{ij} .

2. Definition of a class of spatially-structured random-copying models

We consider a network of N nodes (or sites) labelled by the integers $i = 1, 2, 3, \dots, N$. Site i hosts a population of M individuals, each of which is a member of one of a number of species. The dynamics is specified in terms of a set of copying rates μ_{ij} . Each corresponds to an independent Poisson process: at rate μ_{ij} one of the M individuals on site j is copied, and the offspring individual replaces a randomly-chosen individual on site i . See Figure 1. Note that the case $i = j$ is legitimate, and indeed the rates μ_{ii} will typically be nonzero in any particular instance of this model. We could also consider models where M varies from site to site within the framework presented below: however, this serves only to complicate certain definitions and expressions so we shall restrict the discussion to a spatially-uniform M .

The rates μ_{ij} determine the model's network structure. Elements of the *adjacency matrix* A_{ij} , which indicate if sites i and j are neighbours, are zero if $\mu_{ij} = 0$, and unity otherwise (except the diagonal elements A_{ii} which are always taken to be zero). The *degree* z_i of node i can then be defined as

$$z_i = \sum_j A_{ij}. \quad (2)$$

Although we do not require the adjacency matrix to be symmetric in the first part of this paper, we will for simplicity assume this is the case. (For example, we do not then have to distinguish between in-degree and out-degree). Note, however, that the symmetry $A_{ij} = A_{ji}$ does *not* imply that $\mu_{ij} = \mu_{ji}$.

In a finite system, there is some probability that a species eventually goes extinct—this occurs if all its representatives have been replaced by individuals of other species. Over time, the number of species decreases until only a single one remains. This stationary absorbing state of complete order is referred to variously as *consensus* or *fixation* (in the sociodynamics and genetics literature respectively).

2.1. Voter, invasion and link dynamics

The original voter model dynamics specifies that, in each timestep, a site is first chosen at random, and then a copy taken from one of its neighbours. This constrains the copying rates μ_{ij} to have the form

$$\mu_{ij}^V = \frac{\mu A_{ij}}{z_i} \quad (3)$$

where μ is the rate at which a site is chosen for update. This is, however, only one of a vast range of possible ways to construct a set of copying rates. Also discussed in the literature is the *invasion process*, whereby the parent site j is chosen uniformly at random in each timestep, and the offspring ‘invades’ a randomly chosen neighbour (see, e.g., [9, 10]). This corresponds to the choice

$$\mu_{ij}^I = \frac{\mu A_{ij}}{z_j} . \quad (4)$$

Another possibility is *link dynamics* [9, 10], in which an edge between two nodes is chosen at random, and the copy made in one of the two possible directions along that edge. One way to specify this process is through the rates

$$\mu_{ij}^L = \frac{\mu A_{ij}}{\bar{z}} \quad (5)$$

where \bar{z} is the mean degree within the network. In networks where all nodes have the same degree, $z_i = z_j = \bar{z}$ for all i, j , and these three choices are all equivalent. On heterogeneous networks, these different rules have been found to lead to different scalings of the effective size N_e with the actual network size N . For concreteness, we will mostly consider the above copying schemes. However, we stress that many of our results are quite general, and are not restricted to these specific choices.

2.2. Discrete- and continuous-frequency models

In the original formulation of the voter model, there is one individual per site: $M = 1$. Then, the frequency of a given species on a site can take only one of two values: 0 or 1. To obtain a diffusion equation for the copying dynamics without the need to average over multiple sites (for example, all sites of the same degree, as has been done elsewhere [9, 10, 23, 25–27]), we will take the limit $M \rightarrow \infty$, so that the frequency x_i of a chosen species on a site becomes a continuous variable in the range $[0, 1]$. We will use the terminology *discrete-frequency* and *continuous-frequency* models to refer to the extreme cases $M = 1$ and $M \rightarrow \infty$ respectively.

It is not *a priori* obvious that results obtained for continuous-frequency models will apply to their discrete-frequency counterparts. However, it turns out that for any fixed, finite M , discrete-frequency dynamics are recovered in a limit where the copying rates between sites, μ_{ij} where $i \neq j$, vanish. This fact can be deduced from a convergence theorem for Markov chains due to Möhle [37]. The basic physical idea behind this result is that in the time it takes for an individual to arrive from a neighbouring site,

sufficiently many replacements have been made by sampling from the local population of M individuals that each site is in one of its locally absorbing states—that is, each site hosts only a single species. As a consequence, the frequency of each species on any chosen site is always close to 0 or 1 (although globally it may be somewhere in between) and thus the discrete-frequency dynamics is recovered.

The way in which the $M = 1$ discrete-frequency dynamics is obtained mathematically is summarised in Appendix A. We illustrate here with the concrete examples of the previous subsection. Discrete-frequency versions of voter, invasion or link dynamics can be obtained by taking the parameter μ that appears in (3)–(5) to zero, whilst holding all μ_{ii} at fixed, finite (but otherwise arbitrary) values. To recover the $M = 1$ version of the voter, invasion or link dynamics, with the unit of time defined such that each site is visited on average once per timestep, one must multiply all times by μ/M (see Appendix A).

A key point is that for any finite M , no matter how large, the discrete-frequency dynamics are recovered for sufficiently small μ . That is, any difference between a set of frequencies $\{x_i\}$ in the $M > 1$ model and their values in corresponding realisations of the $M = 1$ model can be made arbitrarily small by reducing μ . In the next section, and the remainder of this work, we will be working in the $M \rightarrow \infty$ limit, such that all frequencies x_i can be regarded as continuous variables. We will assume that results for discrete frequency models can then be obtained by subsequently taking $\mu \rightarrow 0$. In other words, we assume that the limits $M \rightarrow \infty$ and $\mu \rightarrow 0$ commute. We do not know *a priori* that this assumption is valid. In practice, we find that we recover known results for the voter model, or good agreement with simulation, when we perform this procedure (see Section 7).

3. Forward-time formulation as a Fokker-Planck equation

We now derive the Fokker-Planck equation for probability distribution over the N -dimensional space of local frequencies x_i of a chosen species. Although the method used to derive this equation, based on *jump moments*, is standard [38], we will recapitulate it here as it central to the derivation Eq. (1) that follows in Section 5 below. The key point is that the Fokker-Planck equation has the expansion

$$\begin{aligned} \frac{\partial P(\{x_i\}, \tau)}{\partial \tau} = & - \sum_i \frac{\partial}{\partial x_i} a_i^{(1)}(\{x_i\}, \tau) P(\{x_i\}, \tau) + \\ & \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}^{(2)}(\{x_i\}, \tau) P(\{x_i\}, \tau) + \dots \end{aligned} \quad (6)$$

where τ is the time variable and the jump moments are defined as

$$a_{i_1 i_2 \dots i_n}^{(n)}(\{x_i\}, \tau) = \lim_{\delta \tau \rightarrow 0} \frac{\langle [x_{i_1}(\tau + \delta \tau) - x_{i_1}(\tau)] \cdots [x_{i_n}(\tau + \delta \tau) - x_{i_n}(\tau)] \rangle}{\delta \tau}. \quad (7)$$

In calculating the jump moments, the initial coordinates $x_i(\tau)$ are assumed known and fixed, and the angle brackets denote an average over all possible realisations of the

dynamics in the time interval $[\tau, \tau + \delta\tau]$. These averages can be computed from the underlying stochastic dynamical rules.

Within the present class of random-copying models, a local frequency x_i may increase or decrease by an amount $\frac{1}{M}$ in any single update. The continuous-time transition rates for these processes are

$$W\left(x_i \rightarrow x_i + \frac{1}{M}\right) = (1 - x_i) \sum_j \mu_{ij} x_j \quad (8)$$

$$W\left(x_i \rightarrow x_i - \frac{1}{M}\right) = x_i \sum_j \mu_{ij} (1 - x_j) . \quad (9)$$

In the time interval $[\tau, \delta\tau]$, only events involving a single change to one of the x_i contributes to order $\delta\tau$ in the average in (7), for any value of n . Thus, for the first jump moment we have,

$$a_i^{(1)} = \frac{1}{M} \sum_{j \neq i} \mu_{ij} (x_j - x_i) , \quad (10)$$

and for the second moments

$$a_{ii}^{(2)} = \frac{1}{M^2} \sum_j \mu_{ij} [x_i(1 - x_j) + x_j(1 - x_i)] \quad (11)$$

while $a_{ij}^{(2)} = 0$ for $i \neq j$, since two different x_i never change at the same time.

If all μ_{ij} remain fixed as $M \rightarrow \infty$, the second jump moments vanish relative to the first. The resulting Fokker-Planck equation then has only a drift (deterministic) term, and no diffusive component. Both terms are contribute in this limit if one takes μ_{ij} to vanish as $1/M$ for all $j \neq i$, as is standard in population genetics [12, 16]. We thus define rescaled rates m_{ij} and c_i through

$$\mu_{ij} = \begin{cases} \frac{m_{ij}}{M} & i \neq j \\ \frac{c_i}{2} & i = j \end{cases} . \quad (12)$$

By introducing a rescaled time $t = \tau/M^2$ we find as $M \rightarrow \infty$ the Fokker-Planck equation

$$\begin{aligned} \frac{\partial}{\partial t} P(\{x_i\}, t) = & \sum_{\langle i, j \rangle} \left(m_{ij} \frac{\partial}{\partial x_i} - m_{ji} \frac{\partial}{\partial x_j} \right) (x_i - x_j) P(\{x_i\}, t) + \\ & \frac{1}{2} \sum_i c_i \frac{\partial^2}{\partial x_i^2} x_i (1 - x_i) P(\{x_i\}, t) . \end{aligned} \quad (13)$$

The third and higher jump moments are all of order $1/M^3$ and higher, and thus do not contribute in the $M \rightarrow \infty$ limit. Although this Fokker-Planck equation is strictly valid in this limit, it is understood that it serves as an excellent approximation for large but finite M [16]. Our experience suggests that $M \sim 100$ yields results consistent with infinite M (see Section 7.2 for a comparison between finite M simulations and the infinite- M theory).

In principle, to study the ordering dynamics of random-copying processes one would need to solve the full time-dependence of this N -dimensional diffusion. It is therefore

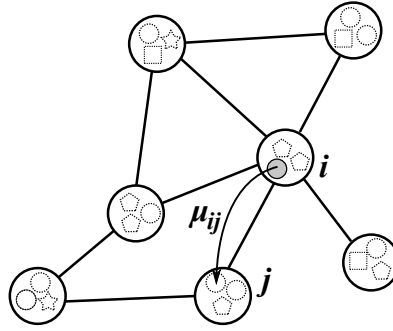


Figure 2. Backward-time formulation of the random-copying dynamics as a random walk process. Shown is the configuration of the system of Figure 1 after the copying event has occurred. Looking backward in time, the tagged individual on site i , shown with a solid fill, has a probability μ_{ij} per unit time of having been copied from site j . The tagged individual thus hops to site j at rate μ_{ij} .

apparent that the reduction to the single coordinate diffusion (1), when it is possible, represents a dramatic reduction in the complexity of the problem.

4. Backward-time formulation as a particle reaction process

As mentioned in the introduction, the copying process defined in Section 2 has an equivalent formulation as a particle reaction dynamics (see e.g. [28,29] for mathematical formulations and [12] for a more detailed account from a statistical physics perspective than the brief outline we give here). This process is identified by looking backward in time at the history of an individual in the current configuration. At some point in the past, this individual was created by the copying process. The rate at which an individual currently on site i was created by copying an individual on site j is μ_{ij}/M , where the factor M appears because μ_{ij} is the total rate at which one of the M individuals on site i is replaced by an offspring arriving from site j . This individual we call the *ancestor* of that in the current configuration. As we look further back in time, the ancestor performs a random walk, hopping from site i to site j at rate μ_{ij} . See Figure 2. We will also refer to these ancestors as *particles*.

One can also consider the ancestry of multiple individuals. These independently perform random walks, as just described. However there is also the possibility that at some point in the history, one ancestor is the direct offspring of one of the others. The rate at which one of a pair of ancestors occupying sites i and j was copied from the other is $(\mu_{ij} + \mu_{ji})/M^2$. This time, the factor M^2 appears because both parent and offspring have a probability of $1/M$ of being involved in the copying process. After this copying event, the two lines of ancestry merge: one individual is the common ancestor of the original pair. This interaction can be viewed as a particle coalescence reaction.

To convert these rates into the system of rescaled time units defined in the previous section, we multiply them by M^2 . We then find that an ancestor hops from site i to j

at rate m_{ij} , and a pair of particles on site i react at rate c_i . The rate for particles on different sites $i \neq j$ to react vanishes under this rescaling.

Within this backward-time picture, it is important to realise that if one starts with an arbitrary pair of individuals in a finite system and looks far enough back in time the ancestor of one individual will at some point be the direct descendant of the other. That is, any two particles eventually react, leading to an overall reduction in the number of particles over time (i.e., fewer ancestors who contribute descendants to the present-day population). The steady state of the particle reaction process thus comprises a single common ancestor of any set of individuals chosen from the present-time population. If there is a single ancestor common to the *whole* population, then it follows that all individuals at the present time are of the same species—i.e., that complete order has been reached at the present time. Thus the stationary absorbing state of full order in the forward-time dynamics corresponds directly to the stationary state of a single particle in the backward-time picture.

In the remainder of this work we need only to consider the backward-time dynamics starting from a single pair of particles. Two key properties of this reaction-diffusion process are the probability distributions $Q_k(t)$ that a single particle on site i at time $t = 0$ occupies site k at time t and $Q_{k\ell}(t)$ that a pair of particles occupy sites k and ℓ . Note that in the reaction dynamics, we take increasing time to mean further back into the past, with $t = 0$ indicating the present time. When talking about a particle pair, we will consider the particles to be distinguishable, so that the configuration $k\ell$ is not the same as ℓk if $k \neq \ell$. These conventions make writing down the master equations for these probability distributions slightly more straightforward.

The master equation governing the single-particle distribution function is

$$\frac{d}{dt}Q_k(t) = \sum_{i \neq k} [Q_i(t)m_{ik} - Q_k(t)m_{ki}] \equiv - \sum_i Q_i(t)M_{ik}^{(1)}. \quad (14)$$

Here we have introduced the matrix

$$M_{ik}^{(1)} = \begin{cases} -m_{ik} & i \neq k \\ \sum_j m_{ij} & i = k \end{cases}, \quad (15)$$

where the sign of the entries has been chosen so that the eigenvalues of $M^{(1)}$ all have nonnegative real parts. Assuming that the network is such that the stationary distribution $Q_k = \lim_{t \rightarrow \infty} Q_k(t)$ is unique, we have at late times

$$Q_k(t) \sim Q_k + au_k e^{-\lambda_r^{(1)}t} \quad (16)$$

where $\lambda_r^{(1)}$ is the eigenvalue of $M^{(1)}$ with the smallest, nonzero real part.

Meanwhile, the joint probability distribution for a pair of unreacted particles is governed by the master equation

$$\frac{d}{dt}Q_{k\ell}(t) = \sum_i Q_{i\ell}(t)m_{ik} + \sum_j Q_{kj}(t)m_{j\ell} - Q_{k\ell}(t)c_k\delta_{k\ell}. \quad (17)$$

The first two terms in this equation give the contributions to this probability from particles hopping into, or out of, sites k and ℓ . The final term arises from the decrease

in the probability that the two particles have not yet reacted due to the coalescence reaction. We can write this equation more compactly as

$$\frac{d}{dt}Q_{k\ell}(t) = -\sum_{ij}Q_{ij}(t)M_{ij,k\ell}^{(2)} \quad (18)$$

where the matrix $M_{ij,k\ell}^{(2)}$ is

$$M_{ij,k\ell}^{(2)} = -m_{ik}\delta_{j\ell} - m_{j\ell}\delta_{ik} + c_i\delta_{ijk\ell} . \quad (19)$$

Here we use $\delta_{ijk\ell}$ as a shorthand for $\delta_{ij}\delta_{jk}\delta_{k\ell}$, i.e., the object that is zero unless all four indices are the same, in which case it equals unity.

On a finite network where the stationary distribution for a single particle is unique, the two particles will eventually react. Therefore $Q_{k\ell}(t)$ vanishes as $t \rightarrow \infty$. The matrix $M^{(2)}$ has eigenvalues whose real parts are all positive. The smallest eigenvalue $\lambda_q^{(2)}$ is strictly positive, and is equal to the asymptotic rate at which a pair of unreacted particles coalesce. The associated eigenvector we will denote $Q_{k\ell}$ and normalise so that $\sum_{k\ell}Q_{k\ell} = 1$. This is the *quasi-stationary distribution* of the two-particle system. That is, if we condition on the two particles never having reacted, the probability they are found on sites k and ℓ approach the stationary value $Q_{k\ell}$ [39]. The rate of relaxation to the quasi-stationary state is given by the real part of the second-smallest eigenvalue $\lambda_r^{(2)}$. That is, at late times

$$Q_{k\ell}(t) \sim Q_{k\ell}e^{-\lambda_q^{(2)}t} + bv_{k\ell}e^{-\lambda_r^{(2)}t} . \quad (20)$$

As stated in the introduction, and to be demonstrated below, it is a separation of timescales between the relaxation to and decay of the quasi-stationary state, controlled by $\lambda_r^{(2)}$ and $\lambda_q^{(2)}$ respectively, that allows for the reduction of the full Fokker-Planck equation (13) to the single-coordinate diffusion of Eq. (1).

5. Exact reduction to a single-coordinate diffusion equation

Now that the forward- and backward-time formulations of the model defined in Section 2 have been set up, we may derive the single-coordinate diffusion (1) and establish the condition for its validity. This amounts to identifying the quantity ξ whose ensemble-average is conserved by the dynamics, and calculating its second jump moment. The procedure is not hard, but it does involve several manipulations that require explanation.

The key quantity ξ has been identified a number of times before (see e.g., [9, 10, 20, 24]). It is

$$\xi(t) = \sum_i Q_i x_i(t) , \quad (21)$$

where we recall that Q_i is the stationary distribution of the single-particle random walk, and x_i is the frequency of the species of interest on site i . To show that the diffusion equation (1) has no deterministic term, we calculate the first jump moment

$$a^{(1)}(\xi, t) = \sum_i Q_i \lim_{\delta t \rightarrow 0} \frac{\langle x_i(t + \delta t) - x_i(t) \rangle}{\delta t} \quad (22)$$

$$= \sum_i Q_i \frac{d}{dt} \langle x_i(t) \rangle = \frac{d}{dt} \langle \xi(t) \rangle . \quad (23)$$

By multiplying (13) by x_i and integrating over all the x coordinates (see [24, 40]), we find

$$\frac{d}{dt} \langle x_i(t) \rangle = \sum_j [m_{ij} \langle x_j(t) \rangle - m_{ji} \langle x_i(t) \rangle] = - \sum_j M_{ij}^{(1)} \langle x_j(t) \rangle \quad (24)$$

where $M^{(1)}$ is as defined in Eq. (15). Hence,

$$a^{(1)}(\xi, t) = - \sum_j \left(\sum_i Q_i M_{ij}^{(1)} \right) \langle x_j(t) \rangle = 0 \quad (25)$$

where the second equality follows because Q_i is the stationary distribution of the single-particle random walk dynamics. In other words, it is the left eigenvector of the matrix $M^{(1)}$ with eigenvalue zero. Thus the first jump moment of ξ vanishes, as required.

We now turn to the second jump moment of ξ . For brevity, we define $x_i \equiv x_i(t)$ and $x'_i \equiv x_i(t + \delta t)$. Then,

$$a^{(2)}(\xi, t) = \sum_{ij} Q_i Q_j \lim_{\delta t \rightarrow 0} \frac{\langle [x'_i - x_i][x'_j - x_j] \rangle}{\delta t} \quad (26)$$

$$= \sum_{ij} Q_i Q_j \lim_{\delta t \rightarrow 0} \frac{\langle x'_i x'_j \rangle - x_i x_j}{\delta t} - 2 \sum_{ij} Q_i Q_j \lim_{\delta t \rightarrow 0} \frac{x_i (\langle x'_j \rangle - x_j)}{\delta t} \quad (27)$$

$$= \sum_{ij} Q_i Q_j \frac{d}{dt} \langle x_i(t) x_j(t) \rangle - 2 \xi \frac{d}{dt} \langle \xi(t) \rangle . \quad (28)$$

Notice that we take all the x_i variables to have known values at time t . There is a slight complication here: only the value of ξ is known at time t , not the individual x_i . However, we will shortly find that at late times $a^{(2)}(\xi)$ depends only on ξ , and hence the resulting expression will apply no matter what the actual individual values of x_i are in any given realisation of the dynamics.

As we have shown above, $\langle \xi(t) \rangle$ has a zero time derivative, so only the first term in the previous expression contributes to the second jump moment. To make progress, it is useful to introduce the quantity

$$P_{ij}(t) = \langle x_i(1 - x_j) \rangle + \langle (1 - x_i)x_j \rangle . \quad (29)$$

This is the probability that two individuals randomly selected from sites i and j are of different species. On a finite network, this quantity asymptotically decays to zero, and in fact, its time evolution is governed by the equation

$$\frac{d}{dt} P_{ij}(t) = -M_{ij,kl}^{(2)} P_{kl}(t) \quad (30)$$

where $M^{(2)}$ is the transition matrix for the two-particle reaction process defined by Eq. (19). One way to see this is to calculate the time derivative of $\langle x_i x_j \rangle$ by multiplying (13) by $x_i x_j$ and integrating, as described previously. A more direct way is to consider events that cause the probability that two individuals are of different species to change in a time interval $[t, t + \delta t]$ (see, e.g., [41]). The rate at which $P_{ij}(t)$ changes due to a

parent on site k leaving an offspring on site i is $m_{ik}[P_{kj}(t) - P_{ij}(t)]$. Meanwhile, if both individuals are on the same site i , there is a probability c_i per unit time that one of them replaces the other (see Section 4), an event that causes the probability that the two individuals are distinct to decrease at rate $c_i P_{ii}(t)$. Summing over all events that change P_{ij} we find

$$\frac{d}{dt}P_{ij}(t) = \sum_{k \neq i} m_{ik}[P_{kj}(t) - P_{ij}(t)] + \sum_{\ell \neq j} m_{j\ell}[P_{i\ell}(t) - P_{ij}(t)] - c_i \delta_{ij} P_{ii}(t) \quad (31)$$

which is equivalent to (30).

Since the time evolution of $P_{ij}(t)$ and the particle pair distribution function $Q_{k\ell}(t)$ discussed in Section 4 are both determined by the same matrix $M^{(2)}$, $P_{ij}(t)$ has a late-time expansion of the same general form as that given by Eq. (20) for $Q_{k\ell}(t)$. That is,

$$P_{ij}(t) \sim P_{ij} e^{-\lambda_q^{(2)} t} + c w_{ij} e^{-\lambda_r^{(2)} t} \quad (32)$$

Now let us suppose there is some limit, for example $N \rightarrow \infty$, in which

$$\frac{\lambda_q^{(2)}}{\text{Re} \lambda_r^{(2)}} \rightarrow 0. \quad (33)$$

Then, under a further rescaling of time $t' = \lambda_q^{(2)} t$, we have in this limit at any $t' > 0$

$$P_{ij}(t') = P_{ij} e^{-t'} \implies \frac{d}{dt'} P_{ij}(t') = -P_{ij}(t'). \quad (34)$$

In other words, at times $t' > 0$, the system is assumed to be in the quasi-stationary state. To avoid too many different time units, we shall remain in the unscaled time t and assume that the quasi-stationarity condition (33) is satisfied, and thus that $\frac{d}{dt}P_{ij}(t) = -\lambda_q^{(2)} P_{ij}(t)$ when $t > 0$.

To complete the derivation of the second jump moment $a^{(2)}(\xi)$, we differentiate both sides of (29) with respect to time t , rearrange and substitute into (28). We find

$$a^{(2)}(\xi, t) = \frac{d}{dt} \langle \xi(t) \rangle - \frac{1}{2} \sum_{ij} Q_i Q_j \frac{d}{dt} P_{ij}(t) \quad (35)$$

$$= \frac{\lambda_q^{(2)}}{2} \sum_{ij} Q_i Q_j P_{ij}(t) \quad (36)$$

$$= \frac{\lambda_q^{(2)}}{2} \sum_{ij} Q_i Q_j (x_i + x_j - 2x_i x_j) = \lambda_q^{(2)} \xi (1 - \xi). \quad (37)$$

To arrive at the second line we used the fact that the average of ξ has zero time derivative, that (33) holds and $t > 0$. The derivation is completed using the definition of $P_{ij}(t)$, Eq. (29), and of ξ , Eq. (21). As advertised, this second jump moment is a function only of the collective coordinate ξ . Using the expansion (6) we find that the motion of the collective coordinate ξ is governed by the Fokker-Planck equation

$$\frac{\partial}{\partial t} P(\xi, t) = \frac{\lambda_q^{(2)}}{2} \frac{\partial^2}{\partial \xi^2} \xi (1 - \xi) P(\xi, t). \quad (38)$$

at any time $t > 0$ in a limit where the condition (33) is satisfied. This is Eq. (1) with an effective size

$$N_e = \frac{2m}{\lambda_q^{(2)}} \quad (39)$$

where we recall m is the baseline rate of the copying process. This concludes the derivation of the main result of this work. In the remaining sections we focus on strategies for determining when the key condition (33) is satisfied, and for estimating the single parameter $\lambda_q^{(2)}$ (or equivalently, N_e) that depends on the network structure and choice of copying dynamics when the diffusion (1) or (38) is valid.

We conclude this section with a couple of remarks on the derivation of (38).

Behaviour at time $t = 0$ At very early times, and technically only at time $t = 0$ in the limit where (33) holds, there are additional contributions to the second jump moment. Whether looking forwards or backwards in time, their net effect is to project the initial condition onto the quasi-stationary state. To be more precise, if the leading eigenvectors of the matrix $M^{(2)}$ are normalised so that

$$\sum_{ij} P_{ij} Q_{ij} = 1, \quad (40)$$

the probability that two particles initially occupying sites i and j have not reacted by time t and occupy sites k and ℓ is

$$D(ij|k\ell; t) = P_{ij} Q_{k\ell} e^{-\lambda_q^{(2)} t} \quad (41)$$

if all but the quasi-stationary modes are neglected at times $t > 0$. By taking the limit $t \rightarrow 0$ from above, and summing over all k and ℓ , we find that P_{ij} is the probability that this particle pair does not react before the quasi-stationary state is reached.

This immediate collapse onto the quasi-stationary state has implications on the initial condition for the diffusion equation (1), given an initial set of frequencies $x_i(0)$. It is possible to calculate the variance in ξ at the onset of quasi-stationarity by appealing to (29). In the simple, solvable examples discussed in Section 7 below where we find (1) to be valid (Wright's island model and the bipartite network), we find that for any given initial combination of $x_i(0)$ a delta-function initial condition $P(\xi, 0) = \delta(\xi - \sum_i Q_i x_i[0])$ is appropriate. Whether this is generally the case is a question we leave open for future work.

Diffusion of other collective coordinates It turns out that any collective coordinate of the form

$$\chi(t) = \sum_i \omega_i x_i(t) \quad (42)$$

where the ω_i are an arbitrary set of positive weights that sum to unity, satisfies the diffusion equation (1) as long as, in addition to (33) the condition

$$\frac{\lambda_q^{(2)}}{\text{Re}\lambda_r^{(1)}} \rightarrow 0 \quad (43)$$

is also satisfied.

This can be seen by repeating the above derivation of the jump moments with this more general set of weights. Ultimately, one finds

$$a^{(1)}(\chi, t) = \frac{d}{dt} \langle \chi(t) \rangle \quad (44)$$

$$a^{(2)}(\chi, t) = (1 - 2\chi) \frac{d}{dt} \langle \chi(t) \rangle - \frac{1}{2} \sum_{ij} \omega_i \omega_j \frac{d}{dt} P_{ij}(t) . \quad (45)$$

The leading decay modes of the first jump moment is governed by the eigenvalue $\lambda_r^{(1)}$, while the modes corresponding to $\lambda_q^{(2)}$, $\lambda_r^{(2)}$ and $\lambda_r^{(1)}$ contribute to the second. Thus we require both (33) and (43) in order to obtain a vanishing first jump moment, and a time-dependent second jump moment of the form

$$a^{(2)}(\chi, t) = \lambda_q^{(2)} \chi(1 - \chi) \quad (46)$$

at times $t > 0$.

Although all collective coordinates of the form (42) have the same diffusion equation—and crucially, the same effective size N_e —when these conditions hold, we prefer the definition (21) because its first jump moment vanishes at all times, not just in the quasi-stationary state, and only the two-particle decay modes contribute to the second, as opposed to a mixture of the one- and two-particle modes. As a consequence, the condition for validity of the diffusion (21) is slightly weaker and cleaner, as it makes reference only to the two leading decay modes within the two-particle sector. Furthermore, the effect of the dynamics at time $t = 0$ before the onset of the quasi-stationary state is harder to predict when the first jump moment does not vanish identically.

6. Methods of analysis for models with time-reversal symmetry

So far in this work, we have shown that if there is a separation of timescales in random-copying models on networks between the rate of relaxation to the quasi-stationary state and the lifetime of that quasi-stationary state, the exactly solvable diffusion equation (1) correctly describes the ordering dynamics at all but the earliest times. To demonstrate the validity of (1), and compute the crucial parameter N_e , one must in principle find the two smallest eigenvalues of the two-particle reaction matrix $M^{(2)}$ and verify the condition (33). Whilst this represents a major simplification of the full dynamics, there is of course no simple, general expression for these eigenvalues for an arbitrary set of transition rates m_{ij} and c_i .

In this section, we describe simplifications that arise in models where the single-particle hopping process exhibits time-reversal symmetry (detailed balance). In particular, this allows one to relate the relaxation rate to the two-particle quasi-stationary state, $\lambda_r^{(2)}$, to the relaxation rate to the single-particle stationary state $\lambda_r^{(1)}$. This leads to a weaker, but more easily verified, condition for the validity of the diffusion (1). Furthermore, the symmetry allows one to develop a systematic variational method

to bound the asymptotic reaction rate $\lambda_q^{(2)}$ from above, or equivalently, the effective size N_e from below. We develop these methods in this section, and apply them to specific combinations of network structure and update rules in the next.

6.1. Time-reversal symmetry and its consequences

The stationary state of the single-particle random walk is time-reversal symmetric if the detailed balance condition

$$Q_i m_{ij} = Q_j m_{ji} \quad \forall i, j \quad (47)$$

is satisfied. Although this may seem an overly restrictive constraint on the rates m_{ij} , it applies, for example, if they have the general form $m_{ij} = m A_{ij} \alpha_i \beta_j$, where A_{ij} is the adjacency matrix and α_i and β_j are arbitrary positive site-dependent quantities. The three sets of rules discussed in Section 2.1 fall into this category, and so the methods we derive here will be applicable to these cases of interest.

The relation (47) implies a corresponding symmetry in the matrix $M^{(2)}$:

$$Q_i Q_j M_{ij,kl}^{(2)} = Q_k Q_\ell M_{kl,ij}^{(2)}. \quad (48)$$

As a consequence of this symmetry, all eigenvalues λ of $M^{(2)}$ are real, and for any right eigenvector ϕ_{ij}^λ the corresponding left eigenvector is $\psi_{ij}^\lambda \propto Q_i Q_j \phi_{ij}^\lambda$ [42]. The eigenvectors form an orthonormal set, i.e.,

$$\sum_{ij} \psi_{ij}^\lambda \phi_{ij}^{\lambda'} \propto \sum_{ij} Q_i Q_j \phi_{ij}^\lambda \phi_{ij}^{\lambda'} = 0 \quad \text{if } \lambda \neq \lambda'. \quad (49)$$

In particular, this means that the key quantities P_{ij} and Q_{ij} characterising the quasi-stationary state are related by $Q_{ij} = A Q_i Q_j P_{ij}$, where the constant A is fixed by the normalisation conditions $\sum_{ij} Q_{ij} = \sum_{ij} P_{ij} Q_{ij} = 1$. These properties are crucial in the analysis of the quasi-stationary state that follows.

6.2. Relationship between one- and two-particle relaxation rates

We now show that, as a consequence of (48), $\lambda_r^{(2)} \geq \lambda_r^{(1)}$, and hence that the condition (43) implies the condition (33) on the validity of the diffusion equation (1). The condition (43) may be easier to demonstrate than (33), since the relaxation rate of the single-particle dynamics is in principle more accessible than that for the two-particle dynamics. Indeed there are methods to bound $\lambda_r^{(1)}$ from above [43] which would provide a means to demonstrate that (43) is satisfied, although we will not use them here.

One way to show that $\lambda_r^{(2)} \geq \lambda_r^{(1)}$ is to use standard perturbation theory [42]. We assume we have the full set of eigenvalues and eigenvectors of the matrix $M^{(2)}$ for some combination of m_{ij} and c_i , and then increase the coalescence rate on site x by an amount δc . The perturbation of $M^{(2)}$ is then $\delta M_{ij,kl}^{(2)} = \delta c$ if $i = j = k = \ell = x$, and zero otherwise. Then, for an eigenvalue λ of $M^{(2)}$ we have from first order perturbation theory [42]

$$\delta \lambda = \delta c \frac{\sum_{ijkl} \psi_{ij}^\lambda \delta M_{ij,kl}^{(2)} \phi_{kl}^\lambda}{\sum_{ij} \psi_{ij}^\lambda \phi_{ij}^\lambda} = \delta c \frac{Q_x^2 [\phi_{xx}^\lambda]^2}{\sum_{ij} Q_i Q_j [\phi_{ij}^\lambda]^2}. \quad (50)$$

This is a nonnegative quantity. Hence, in particular, $\lambda_r^{(2)}$ must increase (or remain constant) as reaction rates c_i are increased from zero. When all $c_i = 0$, the reactions have no effect, and the two particles independently evolve according to the single-particle process governed by $M^{(1)}$. Hence, the second smallest eigenvalue of $M^{(2)}$ equals that of $M^{(1)}$ when all $c_i = 0$. Since $\lambda_r^{(2)}$ cannot decrease as the reactions are turned on, we have $\lambda_r^{(2)} \geq \lambda_r^{(1)}$, as claimed.

6.3. Variational approach for bounding the asymptotic coalescence rate

Since the eigenvalue $\lambda_q^{(2)}$ that sets the characteristic timescale of the diffusion (1) is the smallest of a set of real eigenvalues, we can use a variational principle [42] to bound it from above. Let us denote such a bound λ_+ . The utility of this is twofold. First, if one can show that in some limit

$$\frac{\lambda_+}{\lambda_r^{(2)}} \rightarrow 0 \quad \text{or} \quad \frac{\lambda_+}{\lambda_r^{(1)}} \rightarrow 0, \quad (51)$$

then the condition (33) for the validity of (1) follows. Second, the expression for λ_+ can be used as an estimate of the asymptotic reaction rate $\lambda_q^{(2)}$.

One can construct a bound λ_+ from an arbitrary vector ϕ_{ij} through the expression

$$\lambda_q^{(2)} \leq \lambda_+ = \frac{\sum_{ijkl} Q_i Q_j \phi_{ij} M_{ij,kl}^{(2)} \phi_{kl}}{\sum_{ij} Q_i Q_j \phi_{ij}^2}, \quad (52)$$

which is obtained by expanding the vector ϕ_{ij} as a sum over the eigenvectors of the matrix $M^{(2)}$ (see, e.g., [42] for a derivation).

There are two key points here. First, this result does not rely on being able to solve the two-particle reaction process. Any choice of ϕ_{ij} leads to an expression for λ_+ . Even if this choice departs dramatically from the exact quasi-stationary eigenvector P_{ij} , it may still be possible to show that one of the ratios (51) vanishes, in turn assuring validity of (1). The second point is that, in the usual way, one can include variational parameters in the expression for ϕ_{ij} , and minimise λ_+ with respect to these parameters to obtain the best estimate of $\lambda_q^{(2)}$ available within the space of vectors spanned by them. We will encounter explicit examples in the next section, details of which are in the Appendix.

7. Applications

In order to make concrete the ideas and methods developed in the previous two sections, we now discuss their application to specific combinations of copying dynamics and network structures. In addition to checking certain known limits, we will use the fact that our approach does not invoke a coarse-graining over multiple lattice sites to obtain new theoretical results for the effect of correlations on the effective size N_e .

7.1. Wright's island model

A spatially-structured random-copying model that is prominent in the population genetics literature is Wright's island model [44]. In this model, each site receives an offspring from one of the other sites (chosen uniformly) at a constant rate m . That is, $m_{ij} = \frac{m}{N-1}$ for all $i \neq j$. The coalescence rate is also taken to be uniform: $c_i = c$ for all i . The limited spatial structure in this model admits exact calculation of the eigenvalues $\lambda_r^{(1)}$ and $\lambda_q^{(2)}$, and therewith demonstration that (1) is exact in the $N \rightarrow \infty$ limit. Despite the venerability of Wright's island model, we believe that the validity of the diffusion equation (1) has not been shown explicitly before, although the resulting expression for the effective population size N_e is well known (see e.g., [32]; also [45] for a review).

We begin by examining the single-particle dynamics. Since all sites are equivalent, the stationary distribution is uniform: $Q_i = \frac{1}{N}$. The rate of relaxation to this steady state, $\lambda_r^{(1)}$, is the smallest nonzero eigenvalue of the eigenvector $M_{ij}^{(1)} = \frac{m}{N-1}(N\delta_{ij} - 1)$. One can verify by inspection that any N -dimensional vector with a single entry equal to 1, another equal to -1 and the rest zero is an eigenvector of $M^{(1)}$. Since there are $N-1$ linearly-independent vectors of this form, all $N-1$ nonzero eigenvalues are degenerate, and one finds

$$\lambda_r^{(1)} = m \frac{N}{N-1} . \quad (53)$$

We now turn to the two-particle dynamics. It is possible to calculate $\lambda_q^{(2)}$ exactly. One way to do this is with the variational method described above. We note that in the quasi-stationary state, all knowledge of the initial condition is lost. Thus there are only two distinct two-particle configurations in the quasi-stationary state: one in which both particles are on the same site, and one in which they are on distinct sites. To construct the ansatz, we set all diagonal entries $\phi_{ii} = \alpha$ and all off-diagonal entries equal to unity. The overall normalisation of the vector is unimportant, since this cancels in the numerator and denominator of (52). Thus we can without loss of generality fix one of the variational parameters and optimise the bound (52) with respect to those that remain. Since the exact quasi-stationary eigenvector is obtained with a particular value of α , it follows that by minimising the right-hand of (52) with respect to α will give us the exact decay rate of the quasi-stationary state.

In the Appendix, we show that in the quasi-stationary state, α and $\lambda_q^{(2)}$ are given by the solution of the pair of equations

$$\alpha = \frac{2m}{2m + c - \lambda_q^{(2)}} \quad (54)$$

$$\lambda_q^{(2)} = \frac{\alpha c}{N - 1 + \alpha} . \quad (55)$$

Although these equations admit two solutions for $\lambda_q^{(2)}$, only one is positive. This is

$$\lambda_q^{(2)} = \frac{2mN + c(N-1)}{2(N-1)} \left[\sqrt{1 + \frac{8mc(N-1)}{[2mN + c(N-1)]^2}} - 1 \right] . \quad (56)$$

We can now use condition (43) to identify limits in which the diffusion (1) with effective size $N_e = 2m/\lambda_q^{(2)}$ is valid. One is the *fast-migration limit*, $\frac{c}{m} \rightarrow 0$, for which exact results are available on arbitrary network structures [46]. For Wright's island model we have

$$\frac{\lambda_q^{(2)}}{\lambda_r^{(1)}} = \frac{N-1}{N^2} \frac{c}{m} + O\left(\left[\frac{c}{m}\right]^2\right) \quad (57)$$

indicating that in this limit, (1) is valid for any N . In this regime, $\lambda_q^{(2)} \sim c/N$. This is consistent with the general result obtained by other means [46], and thus serves as a basic check of the methods derived here.

For general c and m , we have in the infinite system size limit, $N \rightarrow \infty$, that

$$\frac{\lambda_q^{(2)}}{\lambda_r^{(1)}} = \frac{2c}{2m+c} \frac{1}{N} + O\left(\frac{1}{N^2}\right). \quad (58)$$

Applying the condition (43) we find that the diffusion (1) is valid for any combination of m and c with an effective size $N_e = N(1 + 2\frac{m}{c})$. This is in agreement with the large- N expression for the effective population size quoted in the literature (see, e.g., [32, 45]). To achieve this agreement, however, one needs to match up the choices of time units and copying rates that are used in standard population genetics models: the correct choice has the literature effective size related to our own by $N_e^{(\text{lit})} = N_e/4m$ and the reaction rate $c = 1/2M$ where M is the subpopulation size. In these units, m is the fraction of individuals replaced by immigrants in each generation. What is perhaps more significant is that, as far as we are aware, the applicability of this single effective population size over extended timescales has not been demonstrated explicitly even for this extremely simple model, even though it is routinely assumed.

We remark that in the limit $m \rightarrow 0$ where discrete-frequency models, like the original voter model, are recovered (see section 2.2), we find that $N_e = N$, as expected. Thus at least in this case, the limits of infinite subpopulation size M and vanishing migration rates m indeed commute.

7.2. Bipartite network

The bipartite network, previously examined in [10, 23, 26], is insightful because it is perhaps the simplest network structure for which the three rules of Section 2.1 are not equivalent. Furthermore, as we will see below, degree correlations can be relevant, in the sense that they can lead to departures from results previously obtained by assuming an absence of such correlations [10, 19, 23, 27].

The N nodes of the bipartite network are divided into two groups, A and B , with a number n_A of the nodes in group A , and n_B in group B . The rate m_{ij} at which an individual i from group A is replaced from a copy of an individual j from group B is $m_{ij} = m_{AB}$. At rate m_{BA} , copying takes place in the reverse direction. As in Wright's island model, the rate c is taken to be the same on all nodes. Note that individuals are never copied between two nodes belonging to the same group. See Figure 3.

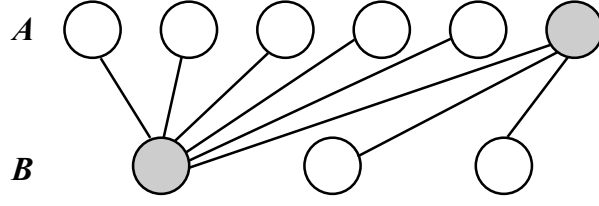


Figure 3. Bipartite network comprising two groups of nodes, A and B . Each site in group A may receive copies from any site in group B with equal probability per unit time, and vice versa. For clarity, the full set of edges has been shown only for the shaded nodes.

Eigenvalue	Degeneracy	Left eigenvector
$\lambda_0 = 0$	1	$(m_{BA}; m_{AB})$
$\lambda_{1A} = n_B m_{AB}$	$n_A - 1$	$([1, -1]; 0)$
$\lambda_{1B} = n_A m_{BA}$	$n_B - 1$	$(0; [1, -1])$
$\lambda_2 = n_A m_{BA} + n_B m_{AB}$	1	$(-n_B; n_A)$

Table 1. Eigenvalues of the single-particle random walk on the bipartite network. $(x; y)$ denotes a N -dimensional vector whose first n_A elements are given by x and the remaining n_B elements by y . The notation $[1, -1]$ indicates a vector of the required dimensionality that has one element equal to 1, another equal to -1 and the rest equal to 0.

To analyse the copying dynamics on this network, we follow the strategy used above for Wright's island model, starting with the single-particle dynamics. One can verify by inspection that the eigenvalues and eigenvectors of the matrix $M^{(1)}$ have the structure presented in Table 1. Note in particular that the elements of the zero left eigenvector give the stationary distribution Q_i (up to a normalisation constant). Hence,

$$Q_i = \begin{cases} \frac{m_{BA}}{n_A m_{BA} + n_B m_{AB}} & 1 \leq i \leq n_A \\ \frac{m_{AB}}{n_A m_{BA} + n_B m_{AB}} & n_A < i \leq n_A + n_B \end{cases} \quad (59)$$

Meanwhile, the second-smallest eigenvalue is

$$\lambda_r^{(1)} = \min\{n_A m_{BA}, n_B m_{AB}\}. \quad (60)$$

To bound the decay rate of the quasi-stationary state, $\lambda_q^{(2)}$, we make use of an extension to the variational ansatz described in the previous section (see Appendix for details). This extended ansatz still takes the quasi-stationary eigenvector $\phi_{ij} = 1$ when two particles occupy different sites $i \neq j$. However, it allows for site-dependent values $\phi_{ii} = \alpha_i$ when both particles occupy the same site. The physical idea behind this ansatz is that the reaction process most strongly affects the probability of finding two particles on the same site, and has a weaker effect elsewhere.

A key quantity in this analysis is the *inverse participation ratio*, which is defined

as

$$\mathcal{I} \equiv \sum_i Q_i^2. \quad (61)$$

Roughly speaking, \mathcal{I}^{-1} measures the number of sites on which the particle may be found in the stationary state (see [47] for further discussion). If \mathcal{I} vanishes as $N \rightarrow \infty$, the stationary distribution is delocalised, and $\lambda_q^{(2)}$ can be shown to be asymptotically bounded from above by

$$\lambda_+^* \sim \sum_i Q_i^2 \frac{2m_i c_i}{2m_i + c_i} \quad (62)$$

where $m_i = \sum_{j \neq i} m_{ij}$ is the total rate at which a particle departs from site i , and c_i is the coalescence rate on site i . Furthermore, if λ_+^* also vanishes with N , the expression given above is the optimal bound on $\lambda_q^{(2)}$ within the space of quasi-stationary eigenvectors admitted by the ansatz. These results are derived in the Appendix. We note also that (62) coincides with an estimate for the effective size that we obtained previously by more approximate means [24].

In the case of the bipartite network, m_i takes on two values: $m_i = n_B m_{AB}$ if i is in group A and $m_i = n_A m_{AB}$ if i is in group B . Recalling that $c_i = c$ is constant, we find

$$\lambda_q^{(2)} \leq \lambda_+^* \sim \frac{2n_A n_B m_{AB} m_{BAC}}{(n_A m_{BA} + n_B m_{AB})^2} \left[\frac{m_{BA}}{2n_B m_{AB} + c} + \frac{m_{AB}}{2n_A m_{BA} + c} \right] \quad (63)$$

if the inverse participation ratio vanishes with N . We can use this bound, along with (60), in (51) to establish cases where the diffusion equation (1) is valid for the bipartite network.

In fact, one can obtain $\lambda_q^{(2)}$ exactly for the bipartite network by following the same procedure as for Wright's island model—that is, by recognising that there are five distinct two-particle configurations in the quasi-stationary state, and identifying each with a variational parameter. If one follows this path, one finds that when \mathcal{I} vanishes, so does the ratio (43), and that $\lambda_q^{(2)} = \lambda_+^*$ asymptotically. That is, we reach exactly the same conclusions with this more complicated approach, so we present only the results obtained using the simpler analysis just described.

Of the three rules discussed in Section 2.1, the voter dynamics and invasion process are the most interesting. We discuss these two cases separately.

7.2.1. Voter dynamics Within voter dynamics, we have

$$m_{AB} = \frac{m}{n_B} \quad \text{and} \quad m_{BA} = \frac{m}{n_A}. \quad (64)$$

The rate of relaxation to the single-particle steady state is simply $\lambda_r^{(1)} = m$. Meanwhile the inverse participation ratio $\mathcal{I} = \frac{N}{4n_A n_B}$, which vanishes as long as both n_A and $n_B = N - n_A$ grow with N as $N \rightarrow \infty$. When this is the case (51) is satisfied, and the asymptotic rate of decay of the quasi-stationary state is

$$\lambda_+^* \sim \frac{2mc}{2m + c} \mathcal{I} = \frac{2mc}{2m + c} \frac{N}{4n_A n_B}. \quad (65)$$

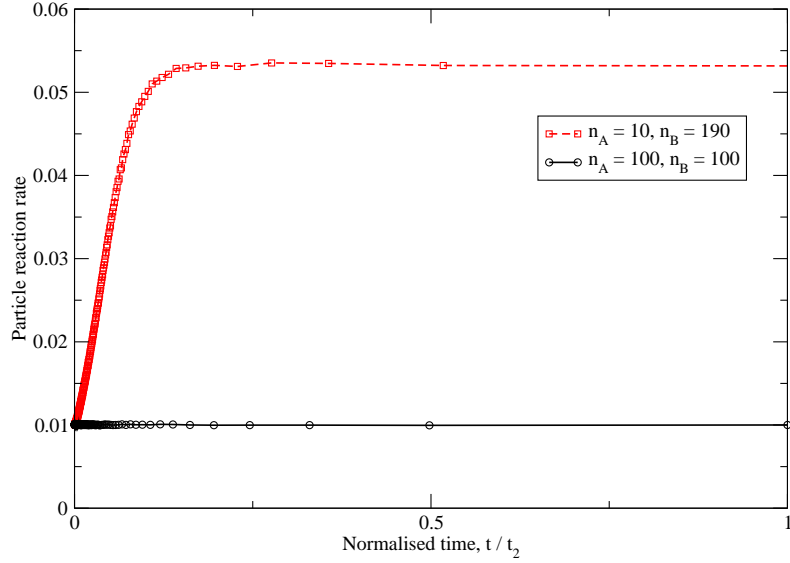


Figure 4. Mean reaction rate per particle pair as a function of time for coalescing random walks on bipartite graphs with hop rates chosen to correspond to the voter model and infinite on-site reaction rate. Symbols show the mean reaction rate within an n -particle sector plotted at the mean time t_n between entering and exiting that sector. These averages were taken over 100000 realisations of the reaction dynamics on any given network structure. Time has been rescaled according to t_2 to allow direct comparison of two different bipartite network structures: one has equal group sizes whilst in the other they differ greatly.

In fact, since in the standard voter dynamics, the total rate of copying into a site, m_i , is the same for every site, λ_+^* is proportional to the inverse participation ratio on any network. We remark that in the limit $\frac{m}{c} \rightarrow 0$ in which the traditional discrete-frequency voter model (see Section 2.2) is expected to be recovered, the result (65) corresponds with that obtained for the voter model on bipartite networks obtained by other means [23, 26].

If, say, n_A remains finite as $N \rightarrow \infty$, the inverse participation ratio no longer vanishes. The stationary distribution is localised within the smaller group. As a consequence, condition (51) is no longer satisfied, suggesting (but not necessarily implying) that there is no separation of timescales in the two-particle dynamics. We test this explicitly with Monte Carlo simulations of the backward-time particle reaction dynamics. Initially, we place one particle on each site, and use the Gillespie algorithm [48] to determine the next particle to hop and the time that it does so. If two particles occupy a site simultaneously, they immediately coalesce: thus we are directly simulating the limit $\frac{m}{c} \rightarrow 0$. In Figure 4, we show an empirical measure of the pairwise reaction rate as a function of time. This is achieved by taking the reciprocal of the mean time spent in the n -particle state (averaged over multiple runs), and dividing by $\binom{n}{2}$ to convert the total reaction rate to a rate per particle pair. As can be seen from Figure 4, in the case where $n_A \ll N$, there is a significant transient in which the reaction rate departs from

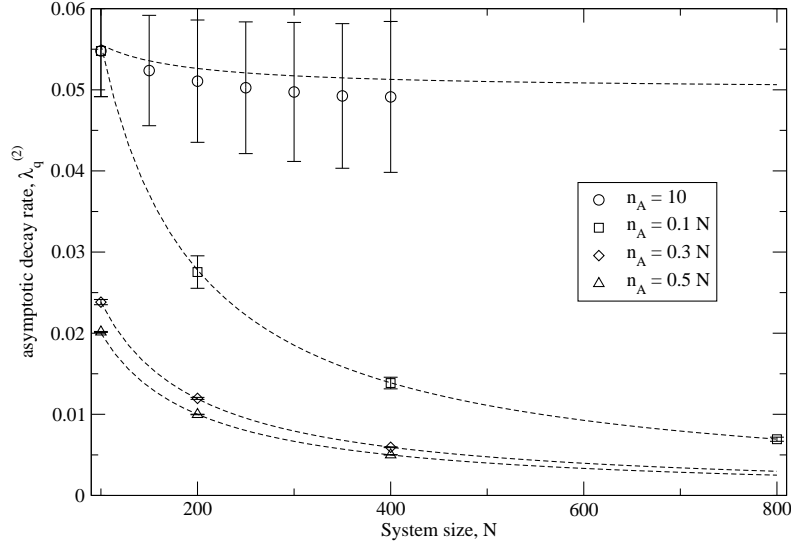


Figure 5. Mean reaction rate per particle pair measured in computer simulations of coalescing random walks on bipartite graphs with hop rates chosen to correspond to the voter model. Circles show the case where one group has a fixed, finite size for different values of the total system size N ; and other symbols cases where both group sizes are proportional to N . Error bars indicate the size of the change in rate over the history of the dynamics (see text for further discussion). The theoretical predictions for the asymptotic decay rate (65) are shown as dashed lines. The discrepancies evident for the case of a fixed group size are expected within the theoretical treatment developed here (see text).

its asymptotic value.

Behaviour in the limit $N \rightarrow \infty$ is revealed by Figure 5. Here, a fit to the asymptotic reaction rate as a function of N is compared for the case $n_A = \text{const}$ versus $n_A \propto N$. This is done by fitting a constant function to data series like those shown in Figure 4. The error bar indicates the standard deviation of the set of rates: this gives an indication of how much transients are contributing to the average. As can be seen, when n_A remains constant as N increases, there is no evidence for a decrease in the contribution of the transient, suggesting that the early-time relaxational dynamics cannot be neglected in the limit $N \rightarrow \infty$, and thus that the diffusion (1) is indeed not valid in this case.

We also examine the effect of intermediate subpopulation size M (see Section 2 for the precise definition) on ordering times. We recall that the limit $M \rightarrow \infty$ is used to obtain the Fokker-Planck equation (13). At finite M it is possible that additional terms may contribute. To investigate this, we performed direct Monte Carlo simulations of the forward-time random-copying dynamics specified in Section 2 with copying rates given by (3). We took the overall rate of copying between sites $\mu = \frac{1}{M}$, so that then $m_{ij} = \frac{A_{ij}}{z_i}$, $m = 1$ and $c_i = c = 2(1 - \frac{1}{M})$. Initially, half of the M individuals on each site was taken to be of the species of interest, so that the initial value of $\xi = \frac{1}{2}$. In the

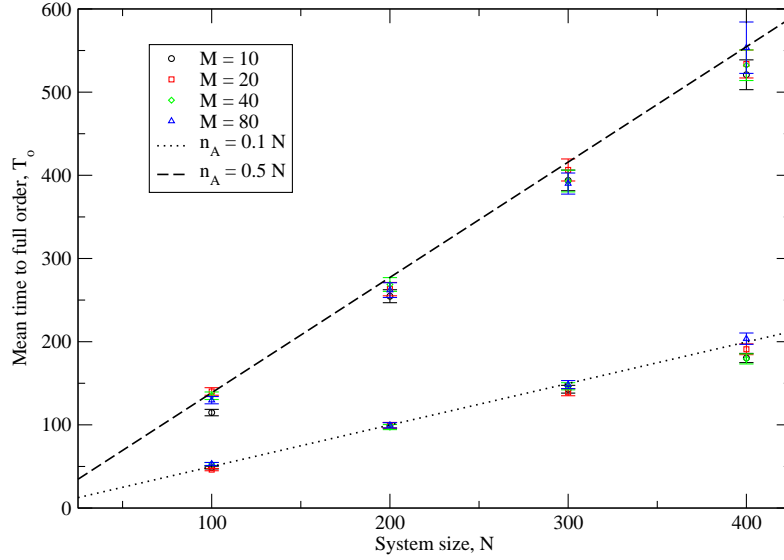


Figure 6. Mean ordering time for the voter model on the bipartite graph with subpopulation sizes $M = 10, 20, 40$ and 80 and group sizes n_A and n_B both proportional to the system size N . Each point was obtained from 500 realisations of the forward-time ordering dynamics from an initial condition comprising equal numbers of two species on each site (the exception is the point with the longest ordering time, $M = 80$, $n_A = n_B = 200$, for which 200 samples were obtained). The dashed lines show predictions from the $M \rightarrow \infty$ theory via Eq. (66).

limit $M \rightarrow \infty$, the effective size appearing in (1) is found from (65) to be

$$N_e = \frac{8n_A n_B}{N}. \quad (66)$$

From a solution of a backward equation corresponding to (1), as described for example in [23, 49], one finds that the mean time T_o for complete order to be reached (i.e. for ξ to reach 0 or 1) from the initial condition $\xi = \frac{1}{2}$ is given by $T_o = N_e \ln 2$. The result of performing many repetitions of the forward-time copying dynamics until order is reached, and averaging the ordering time, is shown in Figure 6 for various M . We see that even with $M = 10$, agreement with the $M \rightarrow \infty$ result is reasonable. Closer inspection of the data reveals that T_o is systematically overestimated in this small M regime; however, already for $M = 80$ the data are statistically indistinguishable from the infinite M prediction. Thus we find that even relatively small subpopulations are reasonably well-described by the continuous frequencies x_i .

7.2.2. Invasion process In the invasion process, the hop rates of Eq. (64) are exchanged. The relaxation rate of the single-particle random walk is

$$\lambda_r^{(1)} = m \times \min \left\{ \frac{n_A}{n_B}, \frac{n_B}{n_A} \right\}. \quad (67)$$

Meanwhile, the inverse participation ratio is given by

$$\mathcal{I} = \frac{n_A^3 + n_B^3}{(n_A^2 + n_B^2)^2} \quad (68)$$

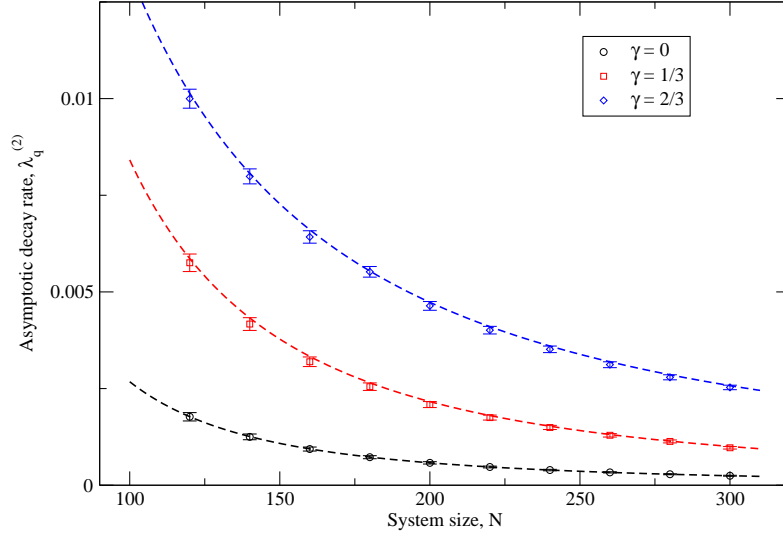


Figure 7. Mean reaction rate per particle pair measured in computer simulations of coalescing random walks on bipartite graphs with hop rates chosen to correspond to the invasion process. The A group had a size $n_A = aN^\gamma$, and the B group $n_B = N - n_A$. The choices of a and γ used were $(a, \gamma) = (10, 0), (5, \frac{1}{3}), (\frac{3}{2}, \frac{2}{3})$. The dashed lines show the prediction from Eq. (69). As in Figure 5, the error bars reflect the size of the contribution from the transient in the numerical estimate of the asymptotic decay rate of the quasi-stationary state.

which vanishes as $N \rightarrow \infty$ for any combination of n_A and n_B , as can be verified by taking $n_A \sim aN^\gamma$ with $0 \leq \gamma \leq 1$ and $n_B = N - n_A$. Then, (63) is asymptotically optimal. Using (51), we then find, again for any combination of n_A and n_B , that the single-coordinate diffusion (1) applies for the invasion process on the bipartite graph.

For $\gamma < 1$ in the above group-size definitions, the estimate for the decay rate of the quasi-stationary state takes the particularly simple large- N form

$$\lambda_+^* \sim \frac{2ma}{N^{2-\gamma}}. \quad (69)$$

We compare the predictions of this formula in the limit $\frac{m}{c} \rightarrow 0$ against simulations of the type discussed above in the context of voter dynamics. As can be seen from Figure 7, numerical measurements of the asymptotic decay rate of the quasi-stationary state agree well with the prediction from Eq. (69). Again this provides evidence that the dynamics of discrete-frequency models is well-represented by the $m \rightarrow 0$ limit of their continuous-frequency counterparts.

We remark that the result (69) *differs* from that previously obtained within an approximation that assumes an absence of degree correlations in the network [10]. Thus the invasion process on the bipartite network is a simple, illustrative example of a case where degree correlations must be taken into account to obtain the correct reduction onto a single-coordinate diffusion of the form (1). Although the correct result was obtained in [26], the separation of timescales required for its validity was assumed, rather than explicitly demonstrated.

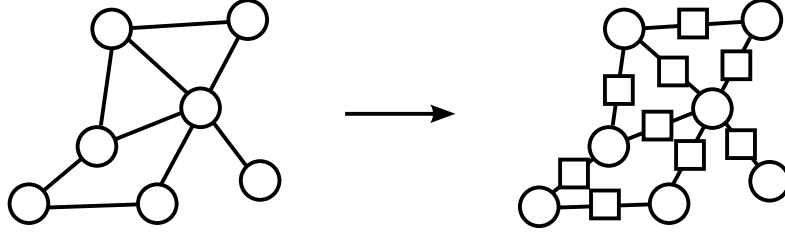


Figure 8. Explosion operation that induces degree anti-correlations in a network. The exploded network (right) is constructed from the template network (left) by inserting a node (shown as squares) at the mid-point of each edge.

7.3. Correlations in random network structures

We now explore this last point in a little more detail by investigating a transformation to a network structure that induces degree anti-correlations. We achieve this by beginning with some template network and then ‘exploding’ it by inserting a node in the middle of each edge. See Figure 8. This way, every node of the template network is then next to a node of degree two: thus high-degree nodes end up next to low-degree nodes.

Clearly, one consequence of this transformation is to increase the size of the network, with a likely corresponding increase in the effective size N_e . Our aim in this final section is to demonstrate that degree anti-correlations lead to a further increase in N_e beyond what is predicted by (62).

As template networks we use Molloy-Reed networks [50] with an approximate power-law degree distribution. These networks are designed to have weak degree correlations. They are constructed by assigning to each node i its desired degree by attaching z_i half-edges to it. Then, these half-edges are formed into complete edges by picking a random pair and joining them together if the two nodes in question are not already connected. Once all half-edges are joined, the probability that nodes i and j are connected is $z_i z_j / \mathcal{K}_1$ where

$$\mathcal{K}_m = \sum_i z_i^m. \quad (70)$$

For this study, we use template networks with the degree distribution

$$p_z = A_\gamma \int_{z-\frac{1}{2}}^{z+\frac{1}{2}} dy y^{-\gamma} \quad (71)$$

where A_γ is a normalisation constant. For large z , this approaches the power-law distribution $p_z \sim z^{-\gamma}$. We shall however mostly be looking at the behaviour on relatively small networks: the purpose of using the above form for p_z is to allow the possibility of highly heterogeneous networks (i.e., a large variance in the degree z). We shall also, for simplicity, restrict ourselves to voter dynamics, i.e., the case where $m_i = m$ on every node, and the limit $\frac{m}{c} \rightarrow 0$. Then, asymptotically, we have from (65) and (61) an upper

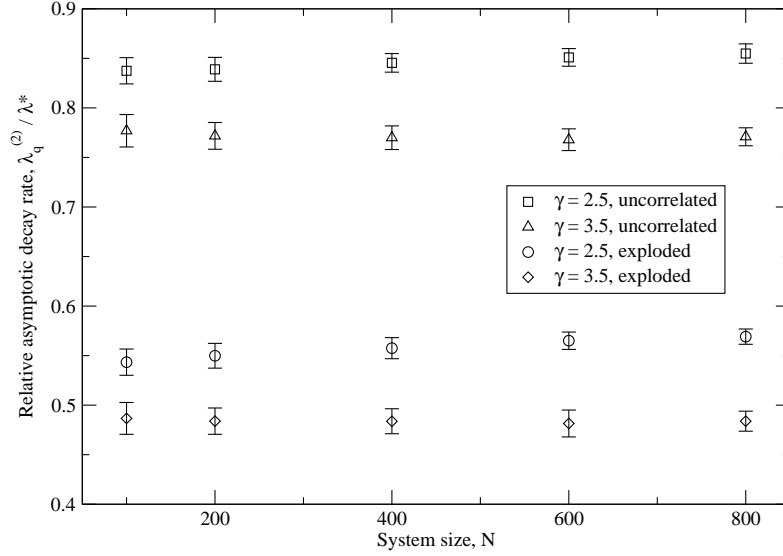


Figure 9. Empirical measurement of the asymptotic decay rate $\lambda_q^{(2)}$ relative to the baseline λ_+^* of Eq. (72) as a function of the size of the uncorrelated template network. Two different values of the exponent γ were used in the power-law degree distribution (71). 50 different random template networks were drawn from this distribution for each γ , and the reaction dynamics repeated 100000 times on both the template and exploded network. In this figure, error bars indicate the standard error of the distribution of reaction rates within the random network ensembles *not* the standard error of the mean.

bound

$$\lambda_+^* \sim 2m\mathcal{I} = 2m\frac{\mathcal{K}_2}{\mathcal{K}_1^2} \quad (72)$$

on $\lambda_q^{(2)}$, since $Q_i = \frac{z_i}{\mathcal{K}_1}$ under voter dynamics [20, 23]. As previously, this bound is valid in the large N limit when \mathcal{I} vanishes with N .

In Figure 9 we show empirical measurements of the asymptotic decay rate of the quasi-stationary state, divided by (72). We used two degree distributions, one with $\gamma = 2.5$, and one with $\gamma = 3.5$. In both cases the minimum degree of any node was 4. For the original template we find deviations from this prediction of 10–20%. Since (72) was obtained by assuming an absence of correlations between particles occupying different sites in the quasi-stationary two-particle state, it would appear that (at least at the network sizes simulated) such correlations exist, despite the fact that degrees of neighbouring nodes are designed to be weakly correlated in Molloy-Reed networks. We also see that the deviations from (72) increase dramatically when anti-correlations are introduced by the explosion procedure discussed above. Although we have not been able to demonstrate analytically that there is a separation of timescales in the dynamics on these networks, numerical data (not shown) suggest that the transient is irrelevant except possibly on the uncorrelated template network with $\gamma = 2.5$.

To understand more quantitatively the decrease of $\lambda_q^{(2)}$ (and hence increase of N_e) under the explosion operation, we employ a refinement of the ansatz used to obtain

(72) that allows for correlations between unreacted particles on neighbouring sites in the quasi-stationary state. Details are presented in the Appendix. The bound on $\lambda_q^{(2)}$ that is ultimately obtained in the $\frac{m}{c} \rightarrow 0$ limit is, for large N ,

$$\lambda_+^\# \sim 2m \frac{\mathcal{K}_2}{\mathcal{K}_1^2} \frac{1 - \frac{\mathcal{K}_2}{\mathcal{S} - \mathcal{T}}}{1 - \frac{\mathcal{S}}{\mathcal{S} - \mathcal{T}} \frac{\mathcal{K}_2}{\mathcal{K}_1^2}}, \quad (73)$$

where

$$\mathcal{S} = \sum_{ij} z_i A_{ij} z_j \quad (74)$$

$$\mathcal{T} = \sum_{ijk} z_i A_{ij} A_{jk} A_{ki}. \quad (75)$$

Here, \mathcal{S} is a measure of how much high-degree nodes tend to be connected to other high-degree nodes: it is related to the *assortativity coefficient* of a network [51]. Meanwhile, \mathcal{T} comprises a sum over triangles in the network, weighted by the total degree of the vertices. It is thus similar in spirit to a *clustering coefficient* [52] in that it vanishes if the network contains no triangles, that is, if locally it has a tree-like structure. The way these quantities interact in (73) is unfortunately not straightforward to interpret. However, the deviation from the baseline (72) will typically be large when degrees of neighbouring nodes are anticorrelated and/or triangles tend to be formed by high-degree nodes.

If nodes are uncorrelated, A_{ij} can be replaced with its ensemble average value in the limit $N \rightarrow \infty$, that is, with the probability $z_i z_j / \mathcal{K}_1$ that nodes i and nodes j are connected [23]. In this case, one finds asymptotically that $\lambda_+^\#$ and λ_+^* converge. On the other hand, the bounds are distinct after the explosion operation. Suppose the template network has given values of \mathcal{K}_1 and \mathcal{K}_2 . Then, after the explosion, there are an additional $\mathcal{K}/2$ additional nodes of degree two. This implies that for the exploded network we have $\mathcal{K}' = 2\mathcal{K}$ and $\mathcal{L}' = \mathcal{L} + 2\mathcal{K}$ (where we use a prime to indicate the exploded network). To compute the statistic \mathcal{S}' , we note that in the exploded graph, each original node i is connected to z_i nodes of degree two. Taking into account that the sum over all neighbouring pairs i, j can be written as twice the sum over all edges, and that each edge is connected to one of the original template nodes, we find that

$$\mathcal{S}' = 4 \sum_i z_i^2 = 4\mathcal{L}. \quad (76)$$

By design, the exploded network contains no triangles, so $\mathcal{T}' = 0$. Using these values in (73) we find, again in the $\frac{m}{c} \rightarrow 0$ limit,

$$\lambda_+^\# \sim 2m\mathcal{I}' \left(\frac{3}{4} - \frac{\mathcal{K}_1}{2\mathcal{K}_2} \right), \quad (77)$$

which we can contrast with the baseline $\lambda_+^* \sim 2m\mathcal{I}'$. Relative to this baseline we see that the asymptotic decay rate of the quasi-stationary state is reduced by a factor of at least $\frac{3}{4}$ as a consequence of the degree anti-correlations in the exploded networks.

In Figure 10 the same data as in Figure 9 are shown, but this time relative to the improved bound (73). We see that in all cases, the improved bound is closer to the

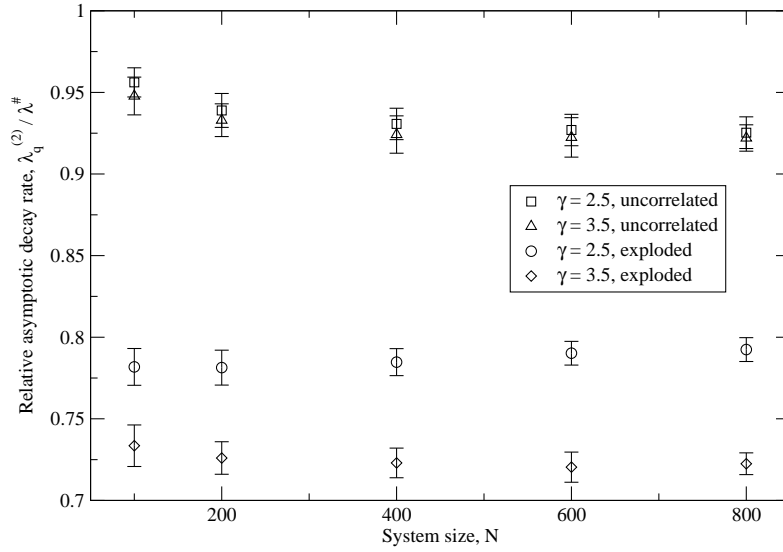


Figure 10. The asymptotic decay rates of Figure 9 rescaled relative to the improved estimate $\lambda^\#$ of Eq. (73). Error bars are to be interpreted as in Figure 9.

measured asymptotic decay rate, although significant discrepancies remain. The reason for this is that the ansatz for the quasi-stationary eigenvector ϕ_{ij} does not contain sufficient structure to capture all the relevant correlations between a pair of unreacted particles. However, the fact that the correction factors shown in Figures 10 and 9 are only weakly dependent on system size N suggests that the correct *scaling* of the quasi-stationary reaction rate with N is captured by the expressions (72) and (73). This is consistent with previous works in which the scaling of N_e with N has been established through fairly simple approximation schemes [9, 10, 19, 23], further work having been necessary to improve on the prefactors [24, 27]. Further improvements demand a more detailed understanding of the structure of the quasi-stationary two-particle state.

8. Discussion and outlook

In this work, we have shown that the onset of order within a large class of random copying processes (including the voter model) on spatially-structured networks can be described exactly in terms of the diffusion of a single collective coordinate ξ . This quantity is a weighted average of a species frequency over nodes of the network. When the reduction onto the single-coordinate diffusion process is possible, the only significant effect of the spatial structure relative to the exactly-solvable ‘mean-field’ diffusion of Equation (1) is to replace the actual network size N with an effective size N_e .

As we have shown in Section 5, a remarkable feature of the analysis is that in order to establish the validity of the single-coordinate diffusion (1), one *only* needs to establish two properties of a two-particle reaction-diffusion process. The first is the lifetime of the quasi-stationary probability distribution for two unreacted particles. The second is a separation of timescales between this quasi-stationary lifetime and that of the relaxation

to quasi-stationarity. In principle, to establish the probability that complete order has set in by a time t , one would need to compute an N -point correlation function, starting, for example, with the full N -dimensional diffusion equation (13). The fact that one can show that the one-dimensional diffusion (1) is valid by appealing only to a two-particle problem represents a considerable simplification in the analysis of voter and other random-copying models.

In Section 7 we demonstrated the separation of timescales within certain exactly-solvable models, namely, Wright’s island model and on the bipartite graph. Although this property has previously been assumed with varying degrees of explicitness within these models [10, 24, 26], we are not aware of this property actually being confirmed for these models before. Furthermore, many existing treatments of random-copying dynamics on heterogeneous networks [9, 10, 23, 25–27] have derived a diffusion approximation by aggregating multiple sites in order to construct continuous spatial variables. Such approaches have the drawback that the network structure gets ‘washed out’ in the coarse-graining. Here we have described an alternative approach whereby a continuous variable is associated with each site, and the network structure is retained. As we have shown in Section 7, anti-correlations between the degree of neighbouring nodes leads in the example of the invasion process on a bipartite network to a scaling of the effective size N_e with network size N that differs from that predicted by an approach that neglects these correlations [10]. For voter dynamics on random networks, the effect of degree correlations appears to be less severe, affecting the prefactor rather than the scaling exponent. Since the inverse participation ratio \mathcal{I} seems to be the quantity that dominates the scaling of N_e with N in these models, the physical origin of the sublinear scaling of N_e with N that can be observed [9–11, 23, 26] lies in the localisation of the quasi-stationary single-particle distribution Q_i in the vicinity of high-degree nodes.

In existing treatments of random-copying dynamics on heterogeneous networks, the methods used to arrive at the diffusion equation (1) have to some extent been *ad-hoc*. Here we have formulated a more systematic treatment based on timescales of the two-particle reaction-diffusion dynamics. The remaining task is to find good methods to estimate these timescales—and thereby demonstrate their separation. In cases where the single-particle reaction dynamics is time-reversal symmetric, we have shown that one can use the single-particle relaxation rate as a proxy for the corresponding quantity in the two-particle sector; we have also demonstrated the utility of a variational approach to bound the effective size N_e from below. While this has allowed the separation of timescales to be demonstrated within very simple models, it was still necessary to resort to simulations to establish whether it was present on random network structures. Development of these methods—perhaps using bounds on the single-particle relaxation dynamics that have been derived elsewhere [43]—would be needed to determine the class of networks and dynamical processes for which (1) is valid.

We remark that (1) is known to apply to the voter model on square lattices in two dimensions and above [28]. In two dimensions, it is found that the effective population size grows logarithmically with system size (relative to the rate at which a particle

relaxes to the stationary uniform distribution). This can be understood as a statement that domains coarsen logarithmically in the two-dimensional voter model [7]. Although the separation of timescales is weak in this case, it is nonetheless present. Thus one would expect the separation of timescales to be observed on random networks with small diameters, on the grounds that a single particle would be expected to explore the system on a faster timescale than would be possible on a two-dimensional square lattice of the same size. It would be interesting to pursue this line of enquiry further.

As mentioned in the introduction, the use of a single effective size N_e to characterise purely stochastic population dynamics in subdivided populations is widespread in population genetics (see e.g., [45] for a review). A fair amount of discussion has centred around the fact that different definitions of N_e are possible, depending on which property of the stochastic dynamics is measured and calibrated against an unstructured (mean-field) population [41]. Whilst it is widely understood that asymptotically many definitions converge onto the lifetime of the quasi-stationary state (also known as the *eigenvalue effective size* [53]), the importance of a separation of timescales between this lifetime and others seems to be less well established. For example, Sjödin *et al* [35], identify the asymptotic rate of particle coalescence as defining ‘the’ effective population size. In these analyses, they examine a separation of timescales in the *definition* of migration and coalescence rates, m_{ij} and c_i . That is, migration rates that vanish with local population size M with differing exponents. This approach potentially misses the possibility—demonstrated here in the examples of Section 7—that one can have a separation of *emergent* timescales that is sufficient to allow a single effective size that characterises almost the entire history of the dynamics. The importance of emergent timescales has perhaps been most clearly recognised by Wakeley and co-workers [33, 34, 36]. In particular, the rapid relaxation of the single-particle dynamics as a condition for a single, well-defined effective size was noted in [36]. However, that study was restricted to a rather special class of *vertex-transitive* networks and dynamical update rules. Furthermore, the irrelevance of *all* timescales other than the relaxation of the quasi-stationary state was implicitly assumed rather than explicitly demonstrated. We believe that the present analysis offers considerable additional insight into the applicability of a one-dimensional diffusion characterised by a single effective population size for a general class of random-copying dynamics on complex networks.

Nevertheless, a number of open questions remain. As we have already noted, the task of calculating the relevant decay rates $\lambda_q^{(2)}$ and $\lambda_r^{(2)}$ is in general challenging (despite being simpler than solving the full N -body dynamics). Although we have shown that certain simplifications arise when the single-particle random walk dynamics are time-reversal symmetric, there is of course a much larger class of processes that do not respect this symmetry: these have nonequilibrium steady states. Here we have not been able to show a result analogous to that derived in Section 6.2, that is, to show that the relaxation rate to two-particle quasi-stationarity exceeds the relaxation rate to a single-particle nonequilibrium steady state, although intuitively one might expect this to be the case. Furthermore, we do not know of any convenient means to bound the

asymptotic reaction rate in these models.

Although we have found that we recover known results for discrete-frequency models, or obtain good agreement with simulations, in the limit $\frac{m}{c} \rightarrow 0$, as we noted in Section 2.2, this correspondence relies on the limits of infinite local subpopulation size and vanishing migration rate to commute. Again, we are not aware of a proof that this should be the case. At the end of Section 5 we commented that the initial condition on the diffusion equation may be modified as the system evolves into the quasi-stationary state: further clarity here would be beneficial. Additionally, although the coordinate ξ is sufficient to track the degree of order present in a system, other aspects of the evolution cannot be determined from this quantity alone—for example, the typical state of the system or correlation length given some value of ξ . It is possible that some of these quantities can be obtained with knowledge of the quasi-stationary eigenvector ϕ_{ij} , this by definition encapsulates two-point correlations. It is unclear whether higher-point correlation functions could be inferred from the two-point functions, or whether these would need to be determined separately.

Finally, we revisit the fundamental property shared by the entire set of random-copying processes discussed here, namely, that all species are equivalent. There is considerable interest in those cases where one species is fitter than another. Although in some sense this is only a minor modification to the dynamics, the mathematics become much more involved. In particular, the backward-time dynamics are more complicated: the reaction rates then depend on the frequencies of different species in the population at the point in history at which the coalescence reaction takes place. It may, however, be possible to apply the results obtained here for the neutral theory in some kind of perturbation theory to study the effects of a small selective advantage. We leave this as one of many possibilities for future work.

Acknowledgments

I thank the RCUK for the support of an Academic Fellowship, Alan McKane and Gareth Baxter for useful discussions during the early development of this work and all those colleagues who have commented on preliminary accounts presented at conferences and research seminars.

Appendix A. Convergence to discrete-frequency models in the limit of slow copying rates

In Section 2.2 it was noted that a convergence theorem due to Möhle [37] implies that if one takes the between-site copying rates $\mu_{ij} \rightarrow 0 \forall i \neq j$ at any fixed subpopulation size M , the discrete-frequency ($M = 1$) dynamics is recovered. In this appendix we provide an informal account of this mathematical result that illustrates how this limit is reached. We refer the reader to [37] for the technical detail.

Our starting point is a *discrete-time* version of the random-copying process. At any

point in time, the state of the system, σ , is specified by the set of numbers n_i which count the number of individuals of the species of interest on site i . In any one timestep, one of these numbers may change by ± 1 . The idea now is to split the matrix A of transition probabilities that appears in the master equation $P_{\sigma'}(t + \delta t) = \sum_{\sigma} A_{\sigma',\sigma} P_{\sigma}(t)$ into two pieces. The first, L , accounts for local interactions: that is, when an individual is replaced by sampling from the same site. The second, I , accounts for interactions between sites. Both are stochastic matrices. If the latter happen a fraction ϵ of the time, then

$$A = (1 - \epsilon)L + \epsilon I. \quad (\text{A.1})$$

The theorem in [37] then states that the vector of probabilities $P(t)$ has the limiting form

$$P(n\delta t) \rightarrow e^{\epsilon n \mathcal{P} I \mathcal{P}} \mathcal{P} P(0) \quad (\text{A.2})$$

as $\epsilon \rightarrow 0$, subject to the existence of the projector

$$\mathcal{P} = \lim_{n \rightarrow \infty} L^n. \quad (\text{A.3})$$

The way to read (A.2) is as follows. The projector \mathcal{P} performs an infinite number of local copying events on every site. If on site i there are initially n_i individuals of the species of interest, then at the end of this process $n_i \rightarrow M$ with probability $\frac{n_i}{M}$, and $n_i \rightarrow 0$ with probability $1 - \frac{n_i}{M}$. Thus the initial condition is projected onto a state in which every site is occupied by individuals of a single species. The exponential term in (A.2) is the generator of a *continuous-time* process. If we take the limit $\delta t \rightarrow 0$, a copy between sites is taken at rate ϵ per timestep. A single copy may leave the system unchanged; or it may change one of the n_i values by ± 1 . In the latter case, all but one of the individuals are of the same species. Then, the projector acts, either to restore the existing value of n_i (with probability $1 - \frac{1}{M}$) or to ‘flip’ it (i.e., from $M \rightarrow 0$ or vice versa) with probability $\frac{1}{M}$.

Putting this together, one finds that the frequency $x_i = \frac{n_i}{M}$ flips from 0 to 1 (or vice versa) due to a copy taken from a neighbouring site j at a rate $\frac{\epsilon}{M}$ multiplied by the probability of copying from j to i in the original discrete-time process. In Section 2, the random-copying process was set up as a continuous-time process in which copies from j arriving at i at a rate μ_{ij} . This is obtained here by putting the transition probabilities encoded in I equal to $\mu_{ij}\delta t$. Then, in the original time units, a copy between sites is taken at rate $\mu'_{ij} = \frac{\epsilon}{M}\mu_{ij}$. Rescaling time such that $t' = \frac{\epsilon}{M}t$, we obtain in the limit $\epsilon \rightarrow 0$ the $M = 1$ (discrete frequency) dynamics from any model with $M > 1$. Note that in the case of the dynamical rules (3)–(5) we can identify ϵ with μ .

Appendix B. Variational bounds on the quasi-stationary decay rate

As stated in the main text, when the backward-time hopping process is time-reversal symmetric, all eigenvalues of the matrix $M^{(2)}$ are real, and hence the smallest eigenvalue

satisfies the bound

$$\lambda_q^{(2)} \leq \lambda_+ = \frac{\sum_{ijkl} Q_i Q_j \phi_{ij} M_{ij,kl}^{(2)} \phi_{kl}}{\sum_{ij} Q_i Q_j \phi_{ij}^2}, \quad (\text{B.1})$$

for any vector ϕ_{ij} . In order to construct bounds on $\lambda_q^{(2)}$ one can use a variational approach. That is, one can include various parameters in ϕ_{ij} , and minimise the right-hand side of (B.1), thereby obtaining the most stringent bound on $\lambda_q^{(2)}$ available within the space of vectors spanned by these parameters.

The key to designing a variational ansatz is to recall that the quasi-stationary distribution of two unreacted particles $Q_{ij} \propto Q_i Q_j \phi_{ij}$. Thus ϕ_{ij} measures the extent to which the probability of finding a pair of particles on sites i and j is affected by the interaction between them. Alternatively, one can infer from the interpretation of $P_{ij} \propto \phi_{ij}$ given at the end of Section 5 that ϕ_{ij} is the probability that two particles starting on sites i and j do not react before the onset of the quasi-stationary state. Either way anticipate that ϕ_{ij} will be close to unity for particles that are far apart, and decrease as their separation decreases.

We make use of two variational ansätze in the main text. In the first it is assumed that ϕ_{ij} differs from unity only when both particles are on the same site. In the second we extend to include the effects of interactions when particles occupy neighbouring sites (but at the expense of restricting to voter-type update rules only).

Ansatz invoking on-site interactions only Restricting the effect of interactions to particles occupying the same site, corresponds to an ansatz for ϕ_{ij} of the form $\phi_{ii} = \alpha_i$ and $\phi_{ij} = 1$ if $i \neq j$. Differentiating (B.1) with respect to α_m , and setting it to zero, we find that

$$\frac{\sum_{kl} M_{mm,kl}^{(2)} \phi_{kl}}{\phi_{mm}} = \frac{\sum_{ijkl} Q_i Q_j \phi_{ij} M_{ij,kl}^{(2)} \phi_{kl}}{\sum_{ij} Q_i Q_j \phi_{ij}^2} = \lambda_+. \quad (\text{B.2})$$

Using the definition of $M^{(2)}$, Eq. (19), we find that

$$\sum_{kl} M_{mm,kl}^{(2)} \phi_{kl} = 2m_i(\alpha_i - 1) + c_i \alpha_i \quad (\text{B.3})$$

where $m_i = \sum_{j \neq i} m_{ij}$, the total exit rate from site i in the random-walk picture. Substituting into (B.2), we find that

$$\alpha_i = \frac{2m_i}{2m_i + c_i - \lambda_+}. \quad (\text{B.4})$$

This is not sufficient to determine the optimal α_i , since the optimal value of λ_+ is not known. To compute this, we need an additional relation between them. This can be obtained as follows. First, we note that (B.2) can be written as

$$Q_i^2 \sum_{kl} M_{ii,kl}^{(2)} \phi_{kl} = \lambda_+ Q_i^2 \phi_{ii}. \quad (\text{B.5})$$

Meanwhile, if we temporarily set $\phi_{ij} = \alpha_{N+1}$ for $i \neq j$, differentiate with respect to this additional α parameter and set to zero, we find analogously that

$$\sum_{i \neq j} Q_i Q_j \sum_{kl} M_{ij,kl}^{(2)} \phi_{kl} = \lambda_+ \sum_{i \neq j} Q_i Q_j \phi_{ij}. \quad (\text{B.6})$$

Summing (B.5) over all i , and adding to this expression, we find that

$$\lambda_+ = \frac{\sum_{ijkl} Q_i Q_j M_{ij,kl}^{(2)} \phi_{kl}}{\sum_{ij} Q_i Q_j \phi_{ij}} = \frac{\sum_i Q_i^2 c_i \phi_{ii}}{\sum_{ij} Q_i Q_j \phi_{ij}} \quad (\text{B.7})$$

where the second equality holds because Q_i is the zero left eigenvector of the matrix $M^{(1)}$. Using the definition (19) and the form of the ansatz for ϕ_{ij} , we ultimately obtain

$$\lambda_+ = \frac{\sum_i Q_i^2 c_i \alpha_i}{1 - \sum_i Q_i^2 (1 - \alpha_i)}. \quad (\text{B.8})$$

The optimal bound on $\lambda_q^{(2)}$ is found by solving the simultaneous set of $N + 1$ equations given by (B.4) and (B.8).

For Wright's island model, discussed in Section 7.1, m_i , c_i and α_i are all site-independent quantities, and $Q_i = \frac{1}{N}$. Making these substitutions in (B.4) and (B.8), we arrive at the pair of equations (54) and (55) studied in the main text.

For more general models, solving (B.4) and (B.8) exactly is difficult. Nevertheless, one can still make some progress here. An alternative way to formulate this problem is to define

$$y(x) = \frac{1}{1 - \sum_i Q_i^2} \sum_i Q_i^2 \frac{2m_i(c_i - x)}{2m_i + c_i - x}. \quad (\text{B.9})$$

Then, the optimal bound λ_+ is given by the value of x for which $y(x) = x$. Differentiating, we find

$$\frac{dy}{dx} = -\frac{1}{1 - \sum_i Q_i^2} \sum_i Q_i^2 \left(\frac{2m_i}{2m_i + c_i - x} \right)^2 \leq 0. \quad (\text{B.10})$$

Hence, the root of $y(x) = x$ must lie at a value of $y(x) \leq y(0)$; hence $\lambda_+ \leq y(0)$. Thus, one has that

$$\lambda_+ \leq \frac{1}{1 - \sum_i Q_i^2} \sum_i Q_i^2 \frac{2m_i c_i}{2m_i + c_i}. \quad (\text{B.11})$$

Note that the right-hand side of this expression is not necessarily the optimal bound.

However, if the inverse participation ratio $\mathcal{I} = \sum_i Q_i^2$ vanishes as $N \rightarrow \infty$, then asymptotically we obtain the simpler expression

$$\lambda_+ \leq \lambda_+^* \sim \sum_i Q_i^2 \frac{2m_i c_i}{2m_i + c_i} \quad (\text{B.12})$$

which was used in the main text. This bound is optimal within the present ansatz for ϕ_{ij} if λ_+ also vanishes as $N \rightarrow \infty$. Then, since

$$\lambda_+ = \frac{\sum_i Q_i^2 (c_i - \lambda_+) \alpha_i}{1 - \sum_i Q_i^2} \quad (\text{B.13})$$

and $\alpha_i \geq \frac{2m_i}{2m_i + c_i}$,

$$\lambda_+ \geq \frac{1}{1 - \sum_i Q_i^2} \sum_i Q_i^2 (c_i - \lambda_+^*) \frac{2m_i}{2m_i + c_i}. \quad (\text{B.14})$$

If both $\mathcal{I} = \sum_i Q_i^2$ and λ_+^* vanish as $N \rightarrow \infty$, this lower bound on the optimal λ_+ asymptotically approaches the upper bound (B.12), and hence the right-hand side of (B.12) is optimal in this case.

Ansatz invoking on-site and nearest-neighbour interactions As noted in the main text, the traditional voter model dynamics is obtained by taking $m_i = m$ and $c_i = c$ followed by the limit $\frac{m}{c} \rightarrow 0$. In the above ansatz, this leads to all $\alpha_i = \frac{2m}{2m+c}$ and $\lambda_+^* \propto \mathcal{I}$. To examine the effects of degree correlations between neighbouring nodes, it is natural to introduce an ansatz for ϕ_{ij} that takes one value when both particles are on the same site, a different value when they are on neighbouring sites, and unity elsewhere. That is,

$$\phi_{ij} = 1 + (\alpha - 1)\delta_{ij} + (\beta - 1)A_{ij} \quad (\text{B.15})$$

where A_{ij} is the adjacency matrix for the network. Recalling that $A_{ij} = 0$ if $i = j$, we see that this expression takes the value α if $i = j$, β if i and j are neighbours, and unity otherwise. For voter dynamics, we have from (3) and (19) that

$$M_{ij,kl}^{(2)} = \left(m\delta_{ik} - \frac{m}{z_i}A_{ik} \right) \delta_{j\ell} + \left(m\delta_{j\ell} - \frac{m}{z_j}A_{j\ell} \right) \delta_{ik} + c\delta_{ijkl} \quad (\text{B.16})$$

where z_i is the degree of node i .

Differentiating the right-hand side of (B.1) with respect to α and setting to zero, we find

$$\sum_{ik\ell} Q_i^2 M_{ii,kl}^{(2)} \phi_{kl} = \lambda_+ \sum_i Q_i^2 \phi_{ii} \quad (\text{B.17})$$

$$\Rightarrow \sum_i Q_i^2 [(2m+c)\alpha - 2m\beta] = \lambda_+ \sum_i Q_i^2 \alpha. \quad (\text{B.18})$$

Hence,

$$\alpha = \frac{2m\beta}{2m+c-\lambda_+}. \quad (\text{B.19})$$

We notice the similarity with (B.4): the on-site correlation ϕ_{ii} is multiplied by the nearest-neighbour correlation β . To obtain an expression for this correlation, we differentiate the right-hand side of (B.1) with respect to β , and set to zero. This yields

$$\sum_{ijk\ell} Q_i A_{ij} Q_j M_{ij,kl}^{(2)} \phi_{kl} = \lambda_+ \sum_{ij} Q_i A_{ij} Q_j \phi_{ij}. \quad (\text{B.20})$$

This expression can be evaluated by using the fact that, for the voter model, the stationary probabilities are $Q_i = z_i/\mathcal{K}_1$. The sum on the right-hand side becomes

$$\frac{1}{\mathcal{K}_1^2} \sum_{ij} z_i A_{ij} z_j \phi_{ij} = \beta \frac{\mathcal{S}}{\mathcal{K}_1^2} \quad (\text{B.21})$$

where \mathcal{K}_n and \mathcal{S} are defined by Equations (70) and (74) respectively. The sum on the left requires a bit more work. One finds it can be written as

$$\begin{aligned} & \frac{2m}{\mathcal{K}_1^2} \left(\sum_{ij} z_i A_{ij} z_j \phi_{ij} - \sum_{ijk} z_i A_{ij} z_j \frac{A_{ik}}{z_i} \phi_{jk} \right) \\ &= \frac{2m}{\mathcal{K}_1^2} \left(\beta \mathcal{S} - \sum_{ijk} z_i A_{ij} A_{jk} [1 + (\alpha - 1)\delta_{ki} + (\beta - 1)A_{ki}] \right) \\ &= \frac{2m}{\mathcal{K}_1^2} (\beta \mathcal{S} - \mathcal{S} - (\alpha - 1)\mathcal{K}_2 - (\beta - 1)\mathcal{T}) \end{aligned} \quad (\text{B.22})$$

in which \mathcal{T} is given by (75). Hence,

$$2m [(\beta - 1)(\mathcal{S} - \mathcal{T}) - (\alpha - 1)\mathcal{K}_2] = \lambda_+ \beta \mathcal{S} \quad (\text{B.23})$$

Substituting (B.19), and rearranging, we finally find

$$\beta = \frac{\mathcal{S} - \mathcal{K}_2 - \mathcal{T}}{\left(1 - \frac{\lambda_+}{2m}\right) \mathcal{S} - \frac{2m}{2m+c-\lambda_+} \mathcal{K}_2 - \mathcal{T}}. \quad (\text{B.24})$$

Similar manipulations allow one to determine that

$$\lambda_+ = \frac{\sum_i z_i^2 \phi_{ii} c}{\sum_{ij} z_i z_j \phi_{ij}} = \frac{\alpha c \mathcal{K}_2}{\mathcal{K}_1^2 + (\alpha - 1)\mathcal{K}_2 + (\beta - 1)\mathcal{S}}. \quad (\text{B.25})$$

When the inverse participation ratio \mathcal{I} vanishes with N , we know that λ_+ does too, since (B.12) bounds it from above, and is proportional to \mathcal{I} under voter dynamics. The upshot of this is that in this case, we can neglect the terms in λ_+ in (B.19) and (B.24). There are further simplifications in the limit $\frac{m}{c} \rightarrow 0$ which is case studied in the main text. To leading order in $\frac{m}{c}$ one finds

$$\alpha = \frac{2m\beta}{c} \quad \text{and} \quad \beta = 1 - \frac{\mathcal{K}_2}{\mathcal{S} - \mathcal{T}}. \quad (\text{B.26})$$

Using these in (B.25), and further taking into account that the ratio $\mathcal{I} = \frac{\mathcal{K}_2}{\mathcal{K}_1^2}$ is assumed to vanish as $N \rightarrow \infty$, we ultimately obtain the expression (73) quoted in the main text.

References

- [1] P. Clifford and A. Sudbury (1973). A model for spatial conflict. *Biometrika* **60** 581.
- [2] R. A. Holley and T. M. Liggett (1975). Ergodic theorems for weakly interacting infinite systems and voter model. *Ann. Probab.* **3** 643.
- [3] J. T. Cox and D. Griffeath (1986). Diffusive clustering in the two-dimensional voter model. *Ann. Probab.* **14** 347.
- [4] T. M. Liggett (1999). *Stochastic interacting systems: contact, voter and exclusion process* (Springer, Berlin).
- [5] J.-M. Drouffe and C. Godrèche (1999). Phase ordering and peristence in a class of stochastic processes interpolating between the ising and voter models. *J. Phys. A: Math. Gen.* **32** 249.
- [6] I. Dornic, H. Chaté, J. Chave and H. Hinrichsen (2001). Critical coarsening without surface tension: The universality class of the voter model. *Phys. Rev. Lett.* **87** 045701.
- [7] C. Castellano, S. Fortunato and V. Loreto (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.* **81** 591.
- [8] A. Baronchelli and R. Pastor-Satorras (2009). Effects of motility on ordering dynamics. *J. Stat. Mech.: Theor. Exp.* L11001.
- [9] C. Castellano (2005). Effect of network topology on the ordering dynamics of voter models. *AIP Conf. Proc.* **779** 114.
- [10] V. Sood, T. Antal and S. Redner (2008). Voter models on heterogeneous networks. *Phys. Rev. E* **77** 41121.
- [11] N. Masuda, N. Gilbert and S. Redner (2010). Heterogeneous voter models. *Preprint* arXiv:1003.0768.
- [12] R. A. Blythe and A. J. McKane (2007). Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech.: Theor. Exp.* P07018.
- [13] S. P. Hubbell (2001). *The unified neutral theory of biodiversity and biogeography* (Princeton University Press).

- [14] M. W. Hahn and R. A. Bentley (2003). Drift as a mechanism for cultural change: an example from baby names. *Proc. Roy. Soc. B* **270** S120.
- [15] D. J. Hruschka, M. H. Christiansen, R. A. Blythe, W. Croft, P. Heggarty, S. S. Mufwene, J. B. Pierrehumbert and S. Poplack (2009). Building social cognitive models of language change. *Trends Cog. Sci.* **13** 464.
- [16] J. F. Crow and M. Kimura (1970). *An introduction to population genetics theory* (Harper and Row, New York, NY).
- [17] F. Slanina and H. Lavička (2003). Analytical results for the Sznajd model of opinion formation. *Eur. Phys. J. B* **35** 279.
- [18] C. Castellano, D. Vilone and A. Vespignani (2003). Incomplete ordering of the voter model on small-world networks. *Europhys. Lett.* **63** 153.
- [19] D. Vilone and C. Castellano (2004). Solution of voter model dynamics on annealed small-world networks. *Phys. Rev E* **69** 016109.
- [20] K. Suchecki, V. M. Eguíluz and M. San Miguel (2005). Conservation laws for the voter model in complex networks. *Europhys. Lett.* **69** 228.
- [21] K. Suchecki, V. M. Eguíluz and M. San Miguel (2005). Voter model dynamics in complex networks: Role of dimensionality, disorder and degree distribution. *Phys. Rev E* **72** 036132.
- [22] C. Castellano, V. Loreto, A. Barrat, F. Cecconi and D. Parisi (2005). Comparison of voter and Glauber ordering dynamics on networks. *Phys. Rev. E* **71** 066107.
- [23] V. Sood and S. Redner (2005). Voter model on heterogeneous graphs. *Phy. Rev. Lett.* **94** 178701.
- [24] G. J. Baxter, R. A. Blythe and A. J. McKane (2008). Fixation and consensus times on a network: A unified approach. *Phy. Rev. Lett.* **101** 258701.
- [25] F. Vazquez and V. M. Eguíluz (2008). Analytical solution of the voter model on uncorrelated networks. *New J. Phys.* **10** 063011.
- [26] C. M. Schneider-Mizell and L. M. Sander (2009). A generalized voter model on complex networks. *J. Stat. Phys.* **136** 59.
- [27] E. Pugliese and C. Castellano (2009). Heterogeneous pair approximation for voter models on networks. *Europhys. Lett.* **85** 58004.
- [28] J. T. Cox (1989). Coalescing random walks and voter model consensus times on the torus in z^d . *Ann. Probab.* **17** 1333.
- [29] M. Notohara (1990). The coalescent and the genealogical process in geographically structured population. *J. Math. Biol.* **29** 59.
- [30] S. Wright (1931). Evolution in Mendelian populations. *Genetics* **16** 97.
- [31] M. Nordborg (1997). Structured coalescent processes on different time scales. *Genetics* **146** 1501.
- [32] J. Wakeley (1999). Nonequilibrium migration in human history. *Genetics* **153** 1863.
- [33] J. Wakeley (2001). The coalescent in an island model of population subdivision with variation among demes. *Theor. Pop. Biol.* **59** 133.
- [34] J. Wakeley and T. Takahashi (2004). The many-demes limit for selection and drift in a subdivided population. *Theor. Pop. Biol.* **66** 83.
- [35] P. Sjödin, I. Kaj, S. Krone, M. Lascoux and M. Nordborg (2005). On the meaning and existence of an effective population size. *Genetics* **169** 1061.
- [36] F. A. Matsen and J. Wakeley (2006). Convergence to the island-model coalescent process in populations with restricted migration. *Genetics* **172** 701.
- [37] M. Möhle (1998). A convergence theorem for markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Probab.* **30** 493.
- [38] H. Risken (1989). *The Fokker-Planck equation* (Springer-Verlag, Berlin).
- [39] J. N. Darroch and E. Seneta (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. App. Probab.* **4** 192.
- [40] G. J. Baxter, R. A. Blythe, W. Croft and A. J. McKane (2006). Utterance selection model of language change. *Phys. Rev. E* **73** 46118.
- [41] M. C. Whitlock and N. H. Barton (1997). The effective size of a subdivided population. *Genetics*

- 146** 427.
- [42] M. C. Pease (1965). *Methods of matrix algebra* (Roberts & Co, Englewood).
 - [43] P. Diaconis and D. Stroock (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36.
 - [44] S. Wright (1943). Isolation by distance. *Genetics* **28** 114.
 - [45] J. Wang and A. Caballero (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82** 212.
 - [46] T. Nagylaki (1980). The strong-migration limit in geographically structured populations. *J. Math. Biol.* **9** 101.
 - [47] F. Wegner (1980). Inverse participation ratio in $2+\epsilon$ dimensions. *Z. Phys. B* **36** 209.
 - [48] D. Gillespie (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.* **22** 403.
 - [49] M. Kimura and T. Ohta (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61** 763.
 - [50] M. Molloy and B. Reed (1995). A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6** 161.
 - [51] M. E. J. Newman (2002). Assortative mixing in networks. *Phy. Rev. Lett.* **89** 208701.
 - [52] D. J. Watts and S. H. Strogatz (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393** 440.
 - [53] W. J. Ewens (2004). *Mathematical Population Genetics: I: Theoretical Introduction* (Springer, New York).